

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 30, 2004, 14:44:25 ; Search time 0.787948 Seconds  
(without alignments)  
2731.623 Million cell updates/sec

Title: US-10-719-385-26  
Perfect score: 36  
Sequence: 1 AGGPGV 6

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 35872929 residues  
Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 150 summaries

Database : A\_Geneseq.23Sep04:.\*  
1: geneseq19808:.\*  
2: geneseq19808:.\*  
3: geneseq20008:.\*  
4: geneseq20018:.\*  
5: geneseq20028:.\*  
6: geneseq20038:.\*  
7: geneseq20038:.\*  
8: geneseq20048:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	36	100.0	6	AD071786	AD071786 BFLP0169
2	36	100.0	166	ABU70740	ABU70740 Human adi
3	36	100.0	407	ADJ69011	ADJ69011 Human hea
4	36	100.0	415	AAU27706	AAU27706 Human ful
5	36	100.0	415	ADJ58205	ADJ58205 Human Cez
6	36	100.0	479	ABB97204	ABB97204 Human hum
7	36	100.0	843	AD125089	AD125089 Human pro
8	36	100.0	843	AD126087	AD126087 Human pro
9	36	100.0	858	ABR47408	ABR47408 Breast ca
10	36	100.0	858	AD37584	AD37584 Human nuc
11	36	100.0	1745	ADP76869	ADP76869 Novel hum
12	36	100.0	1745	AD071781	AD071781 BFLP0169
13	36	100.0	1767	AD071782	AD071782 BFLP0169
14	36	100.0	1883	AD080306	AD080306 Novel hum
15	35	97.2	142	AA979569	AA979569 ACXI C-te
16	35	97.2	1036	ABR82256	ABR82256 Human act
17	35	97.2	1068	AD08157	AD08157 Novel pro
18	33	91.7	642	ABR60188	ABR60188 Drosophi1
19	32	88.9	118	AAW65643	AAW65643 HSV-2 gly
20	32	88.9	20	AAK45552	AAK45552 CRY J I P
21	32	88.9	20	AAK45551	AAK45551 CRY J I P
22	32	88.9	20	AAK82501	AAK82501 CRY J I J
23	32	88.9	20	AAK82500	AAK82500 CRY J I J
24	32	88.9	20	ADD93845	ADD93845 Japanese
25	32	88.9	20	ADD93844	ADD93844 Japanese

26	32	88.9	30	AAK45579	AAK45579 CRY J I P
27	32	88.9	30	AAW44683	AAW44683 T-cell ep
28	32	88.9	33	AD11905	AD11905 Ravenfox
29	32	88.9	43	AAW61570	AAW61570 Human imm
30	32	88.9	48	ABO57709	ABO57709 Human gen
31	32	88.9	53	AAW62764	AAW62764 Human imm
32	32	88.9	54	AAW05836	AAW05836 Mouse M-S
33	32	88.9	54	AAW59595	AAW59595 Mouse Ser
34	32	88.9	54	ADL71374	ADL71374 Mouse ser
35	32	88.9	71	ADK34982	ADK34982 Novel hum
36	32	88.9	74	AAW65187	AAW65187 Prolionib
37	32	88.9	74	ABG29902	ABG29902 Novel hum
38	32	88.9	74	ABM61706	ABM61706 Prolionib
39	32	88.9	88	ADAA57035	ADAA57035 Human sec
40	32	88.9	88	ADAA0890	ADAA0890 Human sec
41	32	88.9	88	ABR47812	ABR47812 Human sec
42	32	88.9	88	ADB91585	ADB91585 Human sec
43	32	88.9	88	ADD37916	ADD37916 Human sec
44	32	88.9	89	AAV38413	AAV38413 Human sec
45	32	88.9	89	AAV38428	AAV38428 Human sec
46	32	88.9	89	ABG03528	ABG03528 Novel hum
47	32	88.9	89	ABG03527	ABG03527 Novel hum
48	32	88.9	91	AAV35308	AAV35308 Chlamydia
49	32	88.9	91	AAV63229	AAV63229 Prolionib
50	32	88.9	91	ABM59748	ABM59748 Prolionib
51	32	88.9	108	AAW04030	AAW04030 Human sec
52	32	88.9	108	ABM63313	ABM63313 Drosophi1
53	32	88.9	114	AD140621	AD140621 Human pur
54	32	88.9	116	ABG19371	ABG19371 Novel hum
55	32	88.9	122	AAU64490	AAU64490 Prolionib
56	32	88.9	122	ABM61009	ABM61009 Prolionib
57	32	88.9	126	AAU47152	AAU47152 Prolionib
58	32	88.9	126	ABM43671	ABM43671 Prolionib
59	32	88.9	135	ABO69120	ABO69120 Pseudomon
60	32	88.9	137	ABG19455	ABG19455 Novel hum
61	32	88.9	149	AAAB63271	AAAB63271 Human bre
62	32	88.9	149	AAAB63375	AAAB63375 Human bre
63	32	88.9	152	ABP58806	ABP58806 Calcium-d
64	32	88.9	171	ABO71893	ABO71893 Pseudomon
65	32	88.9	172	ABG11743	ABG11743 Novel hum
66	32	88.9	192	AAW18350	AAW18350 Prolifera
67	32	88.9	194	ADB64161	ADB64161 Human pro
68	32	88.9	194	ABO72453	ABO72453 Pseudomon
69	32	88.9	196	AAW73911	AAW73911 Mutant E2
70	32	88.9	197	ABG14123	ABG14123 Novel hum
71	32	88.9	208	AAW41818	AAW41818 Human pol
72	32	88.9	216	ADMO5789	ADMO5789 Human pro
73	32	88.9	226	ABO78982	ABO78982 Pseudomon
74	32	88.9	247	AAU31932	AAU31932 Novel hum
75	32	88.9	248	ADG75070	ADG75070 Human hex
76	32	88.9	253	ADMO4911	ADMO4911 Human pro
77	32	88.9	264	ABO71628	ABO71628 Pseudomon
78	32	88.9	266	AAAG39368	AAAG39368 Arabidops
79	32	88.9	266	ADB64960	ADB64960 Human pro
80	32	88.9	272	AAAG42704	AAAG42704 Arabidops
81	32	88.9	275	AAV39302	AAV39302 Spnf prot
82	32	88.9	275	AAW70948	AAW70948 S. spinos
83	32	88.9	275	ABP57683	ABP57683 Saccharop
84	32	88.9	276	AAAG39367	AAAG39367 Arabidops
85	32	88.9	280	ABR55197	ABR55197 Amino aci
86	32	88.9	283	ABG14124	ABG14124 Novel hum
87	32	88.9	284	AAW40032	AAW40032 Human pol
88	32	88.9	288	ABG26229	ABG26229 Novel hum
89	32	88.9	288	AAW29812	AAW29812 Novel hum
90	32	88.9	290	ABO74840	ABO74840 Pseudomon
91	32	88.9	296	ABP97797	ABP97797 Amino aci
92	32	88.9	302	ABR61804	ABR61804 Human JAG
93	32	88.9	302	ABR61734	ABR61734 Human JAG
94	32	88.9	343	ADK34132	ADK34132 Novel hum
95	32	88.9	351	ADB64772	ADB64772 Human pro
96	32	88.9	351	AAW75388	AAW75388 Japanese
97	32	88.9	353	AAW81587	AAW81587 Cedar pol
98	32	88.9	365	AAW73912	AAW73912 Mutant E2

150	32	88.9	581.4	ABG25418
149	32	88.9	511.6	ABG732278
148	32	88.9	486.2	ABG08034
147	32	88.9	479.7	ADP53743
146	32	88.9	476.2	ADP58287
145	32	88.9	462.6	ABU48644
144	32	88.9	448.6	ADL57951
143	32	88.9	446.4	ABBI1000
142	32	88.9	443.3	AAAG69746
141	32	88.9	443.3	AAAG69746
140	32	88.9	443.3	AAAG69746
139	32	88.9	437.8	ADBI19480
138	32	88.9	437.8	ADBI19480
137	32	88.9	437.8	ADBI19480
136	32	88.9	437.8	ADBI19480
135	32	88.9	437.8	ADBI19480
134	32	88.9	437.8	ADBI19480
133	32	88.9	437.8	ADBI19480
132	32	88.9	437.8	ADBI19480
131	32	88.9	437.8	ADBI19480
130	32	88.9	437.8	ADBI19480
129	32	88.9	437.8	ADBI19480
128	32	88.9	437.8	ADBI19480
127	32	88.9	437.8	ADBI19480
126	32	88.9	437.8	ADBI19480
125	32	88.9	437.8	ADBI19480
124	32	88.9	437.8	ADBI19480
123	32	88.9	437.8	ADBI19480
122	32	88.9	437.8	ADBI19480
121	32	88.9	437.8	ADBI19480
120	32	88.9	437.8	ADBI19480
119	32	88.9	437.8	ADBI19480
118	32	88.9	437.8	ADBI19480
117	32	88.9	437.8	ADBI19480
116	32	88.9	437.8	ADBI19480
115	32	88.9	437.8	ADBI19480
114	32	88.9	437.8	ADBI19480
113	32	88.9	437.8	ADBI19480
112	32	88.9	437.8	ADBI19480
111	32	88.9	437.8	ADBI19480
110	32	88.9	437.8	ADBI19480
109	32	88.9	437.8	ADBI19480
108	32	88.9	437.8	ADBI19480
107	32	88.9	437.8	ADBI19480
106	32	88.9	437.8	ADBI19480
105	32	88.9	437.8	ADBI19480
104	32	88.9	437.8	ADBI19480
103	32	88.9	437.8	ADBI19480
102	32	88.9	437.8	ADBI19480
101	32	88.9	437.8	ADBI19480
100	32	88.9	437.8	ADBI19480
99	32	88.9	437.8	ADBI19480
98	32	88.9	437.8	ADBI19480
97	32	88.9	437.8	ADBI19480
96	32	88.9	437.8	ADBI19480
95	32	88.9	437.8	ADBI19480
94	32	88.9	437.8	ADBI19480
93	32	88.9	437.8	ADBI19480
92	32	88.9	437.8	ADBI19480
91	32	88.9	437.8	ADBI19480
90	32	88.9	437.8	ADBI19480
89	32	88.9	437.8	ADBI19480
88	32	88.9	437.8	ADBI19480
87	32	88.9	437.8	ADBI19480
86	32	88.9	437.8	ADBI19480
85	32	88.9	437.8	ADBI19480
84	32	88.9	437.8	ADBI19480
83	32	88.9	437.8	ADBI19480
82	32	88.9	437.8	ADBI19480
81	32	88.9	437.8	ADBI19480
80	32	88.9	437.8	ADBI19480
79	32	88.9	437.8	ADBI19480
78	32	88.9	437.8	ADBI19480
77	32	88.9	437.8	ADBI19480
76	32	88.9	437.8	ADBI19480
75	32	88.9	437.8	ADBI19480
74	32	88.9	437.8	ADBI19480
73	32	88.9	437.8	ADBI19480
72	32	88.9	437.8	ADBI19480
71	32	88.9	437.8	ADBI19480
70	32	88.9	437.8	ADBI19480
69	32	88.9	437.8	ADBI19480
68	32	88.9		

## ALIGNMENTS

RESULT 1	
AD071786	
ID	AD071786 standard; protein; 6 AA.
XX	
AC	AD071786;
XX	
DT	26-AUG-2004 (first entry)
XX	
DE	BFLP0169 related protein, SEQ ID 26.
XX	
KW	Antiinflammatory; Dermatological; Immunosuppressive; human; BFLP0169
KW	kidney; lupus nephritis; autoimmune disorder; lupus.
XX	
OS	Homo sapiens.
XX	
PN	WO2004048521-A2.

XX 10-JUN-2004.  
PD  
XX 21-NOV-2003; 2003WO-US037339.  
PF  
XX 21-NOV-2002; 2002US-0428094P.  
PR  
XX (AMHP ) WYETH.  
PA  
XX O'toole M, Mounts WM, Shojaee N,  
PI  
XX WPI; 2004-441163/41.  
DR  
XX  
XX New isolated BFLP0169 nucleic acids and proteins, useful for diagnosing  
PT or treating autoimmune disorder, e.g. lupus or lupus nephritis.  
XX  
XX  
XX Claim 18; Page; 83pp; English.  
PS  
XX  
XX The present invention relates to human BFLP0169 (ADO71763) and its coding  
CC sequence (ADO71762). BFLP0169 gene expression is increased in kidney  
CC tissue in mice with lupus nephritis. The BFLP0169 sequences are useful  
CC for diagnosing and treating autoimmune disorder, preferably lupus or  
CC lupus nephritis. They are also useful in the generation of antibodies  
CC that bind immunospecifically to the nucleic acids. The present sequence  
CC was used to illustrate the invention. Note: The present sequence was not  
CC shown in the specification, but was derived from information given on  
CC page 13 and ADO71781.  
XX  
XX  
XX Sequence 6 Aa;

Query Match	100.0%	Score 36	DB 8	Length 6
Beer Local Similarity	100.0%	Pred. No. 1	7e+05	
Matches	6	Conservative	0	Mismatches 0
				Indels 0
				Gaps 0
QY	1	AGGPGY	6	
DB	1	AGGPGY	6	

PT		New complex between two interacting proteins in adipocyte cells, useful for identifying selected interacting domains that modulate protein interactions, or for preventing or treating metabolic disorders such as
XX		
XX		
DR	N-PSDB; ACA57284.	
XX		
PI	Légrain P, Davlet L;	
XX		
XX		
AC	WPI; 2003-103412/09.	
XX		
XX		
DT	10-JUN-2003 (first entry)	
DE	Human adipocyte Selected Interacting domain, SID, #371.	
XX		
KM	Human; prey; adipocyte; SID; selected interacting domain; anorectic; antidiabetic; protein-protein interaction; diabetes;	
KW	yeast 2-hybrid assay; metabolic disorder; obesity.	
OS	Homo sapiens.	
PN	WO200286122-A2.	
PD	31-OCT-2002.	
PF	14-MAR-2002; 2002MO-EP003768.	
PR	14-MAR-2001; 2001US-0275734P.	
PA	(HYBR-) HYBRIGENICS.	
ID	ABU70740 standard; protein; 166 AA.	
AC	ABU70740;	
RESULT 2		
ABU70740		

PT obesity or diabetes.  
XX  
XX Claim 6; Page 232; 382pp; English.  
XX  
XX The invention relates to a complex between two interacting proteins in  
CC adipocyte cells, given in the specification. The proteins are identified  
CC by selecting a bait protein from a known adipocyte marker and then  
CC performing a yeast 2-hybrid selection to isolate prey proteins encoded by  
CC members of an adipocyte cDNA library. The proteins are designated SID  
CC (RTM) (selected interacting domains) proteins. Also included are a  
CC polynucleotide encoding a polypeptide in the adipocyte cells, a  
CC recombinant host cell expressing at least one of the interacting  
CC polypeptides of the complex, selecting a modulating compound in adipocyte  
CC cells, a SID (RTM) polypeptide comprising any of the 738 amino acid  
CC sequences given in the specification (including its fragment or variant),  
CC a SID (RTM) polynucleotide comprising any of the 738 nucleotide sequences  
CC given in the specification (including its fragment or variant), a vector  
CC comprising the SID (RTM) polynucleotide, a recombinant host cell  
CC comprising the vector, a protein chip comprising the polypeptides and a  
CC record comprising all or part of the data, listed in the specification.  
CC The complex, polypeptides, polynucleotides and compounds are useful for  
CC preventing or treating metabolic disorders such as obesity or diabetes.  
CC The polynucleotides are useful as probes or primers. The complex is  
CC particularly useful for identifying selected interacting domains (SID  
CC (RTM)) for screening drugs that modulate the protein interaction, thus  
CC exhibiting the therapeutic effect. The present sequence represents a SID  
CC (prey) protein of the invention  
XX  
SQ Sequence 166 AA;  
Query Match 100.0%; Score 36; DB 6; Length 166;  
Best Local Similarity 100.0%; Pred. No. 3.6e+02; Indels 0; Gaps 0;  
Matches 6; Conservative 0; Mismatches 0;  
QY 1 AGGPGCV 6  
Db 39 AGGPGCV 44  
RESULT 3  
ADJ69011  
ID ADJ69011 standard; protein; 407 AA.  
XX  
XX ADJ69011;  
AC  
XX  
XX 06-MAY-2004 (first entry)  
DT  
XX  
XX Human heart mitochondrial protein as a therapeutic target SeqID817.  
DE  
XX  
XX Mitochondrial; human; screening assay; diabetes mellitus;  
KM Huntington's disease; osteoarthritis;  
KM Leber's hereditary optic neuropathy; LHON;  
KM mitochondrial encephalopathy lactic acidosis and stroke; MELAS;  
KM myoclonic epilepsy ragged red fibre syndrome; MERRF; cancer;  
KM neuroprotective; noctropic; antidiabetic; anticonvulsant; antiarthritic;  
KM osteopathic; ophthalmological; cytostatic.  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO2003087768-A2.  
PN  
XX  
XX 23-OCT-2003.  
PD  
XX  
XX 04-APR-2003; 2003WO-US010870.  
PF  
XX  
XX 12-APR-2002; 2002US-0372843P.  
PR 17-JUN-2002; 2002US-038987P.  
PR 20-SEP-2002; 2002US-0412418P.  
XX  
XX (MITO-) MITOKOR.  
PA (BUCK-) BUCK INST AGE RES.  
XX  
XX Ghosh SS, Fahy ED, Zhang B, Gibson BW, Taylor SW, Glenn GM,

PI Warnock DE;  
XX  
XX WPI; 2003-845369/78.  
DR  
XX  
XX Identifying a mitochondrial target for drug screening assays and for  
PT treating diseases associated with altered mitochondrial function,  
PT comprises detecting a modified polypeptide in a sample and correlating  
PT with the disease.  
XX  
XX Claim 1; SEQ ID NO 817; 180pp; English.  
PS  
XX  
XX This invention relates to novel mitochondrial targets that can be used  
CC for therapeutic intervention in treating a disease associated with  
CC altered mitochondrial function. Specifically, it refers to a method for  
CC identifying proteins of the human heart mitochondrial proteome that are  
CC useful for drug screening assays, as well as therapeutic targets. The  
CC present invention describes a method for identifying such proteins that  
CC can be used in the treatment of various diseases associated with altered  
CC mitochondrial function including diabetes mellitus, Huntington's disease,  
CC osteoarthritis, Leber's hereditary optic neuropathy (LHON), mitochondrial  
CC encephalopathy lactic acidosis and stroke (MELAS), myoclonic epilepsy  
CC ragged red fibre syndrome (MERRF) or cancer. Accordingly, these  
CC compositions have neuroprotective, noctropic, antidiabetic,  
CC anticonvulsant, antiarthritic, osteopathic, ophthalmological and  
CC cytostatic activities. This polypeptide sequence is a human heart  
CC mitochondrial protein of the invention.  
XX  
SQ Sequence 407 AA;  
Query Match 100.0%; Score 36; DB 7; Length 407;  
Best Local Similarity 100.0%; Pred. No. 7.7e+02; Indels 0; Gaps 0;  
Matches 6; Conservative 0; Mismatches 0;  
QY 1 AGGPGCV 6  
Db 280 AGGPGCV 285  
RESULT 4  
AAU27706  
ID AAU27706 standard; protein; 415 AA.  
XX  
XX AAU27706;  
AC  
XX  
XX 18-DEC-2001 (first entry)  
DT  
XX  
XX Human full-length polypeptide sequence #31.  
DE  
XX  
XX Mammal; human; rhesus monkey; baker's yeast; fission yeast; Norway rat;  
KM mouse; Chinese hamster; African clawed frog; fruit fly; dog; leukemia;  
KM cancer; lymphoma; neuroblastoma; autoimmune disorder; cell proliferation;  
KM nervous system disorder; inflammatory disorder; cell differentiation;  
KM angiogenesis; stem cell growth factor; activin; inhibin; cartilage; burn;  
KM genetic disorder; bone regeneration; tendon; ligament; tissue repair;  
KM cytostatic; antirheumatic; antiarthritic; vulnery; antiinflammatory;  
KM antibacterial; immunosuppressive; vasotropic; antiparkinsonian;  
KM neuroprotective; osteopathic; antidiabetic; antistimatic; antiallergic;  
KM immunostimulant; analgesic; gene therapy.  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO200164834-A2.  
PN  
XX  
XX 07-SEP-2001.  
PD  
XX  
XX 26-FEB-2001; 2001WO-US004926.  
PF  
XX  
XX 28-FEB-2000; 2000US-00515126.  
PR 18-MAY-2000; 2000US-00577409.  
PR 17-JUN-2000; 2000US-00597707.  
PR 14-JUL-2000; 2000US-00616807.  
PR 19-SEP-2000; 2000US-00664641.  
XX

PA (HYSE-) HYSEQ INC.  
 XX Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;  
 PI Xue AJ, Yang Y, Wehrman T, Wang J, Ma Y, Wang D, Chen R, Xu C;  
 PI Drmanac R;  
 XX WPI; 2001-589862/66.  
 DR N-PEDB; AAS44606.  
 PT Novel polypeptides and nucleic acids obtained from cDNA libraries  
 prepared from various human tissues, for diagnosis, treatment of cancer,  
 PT neurological, inflammatory disorders and for use in arrays for detection.  
 XX  
 PS Claim 10; SEQ ID NO 203; 153bp; English.  
 CC Sequences AAU7676-AU28019 represent full-length polypeptides and contig  
 CC polypeptides of the invention. The proteins and their associated DNA  
 CC sequences are useful for the treatment, diagnosis and prevention of  
 CC various types of disorder in a mammalian subject such as a human, dog,  
 CC monkey, mouse, hamster or rat. The disorders include cancers such as  
 CC leukemia, lymphoma and neuroblastoma, autoimmune disorders such as  
 CC multiple sclerosis, connective tissue disease, rheumatoid arthritis,  
 CC diabetes mellitus, allergic rhinitis, asthma and eczema, nervous system  
 CC disorders such as Parkinson's disease, Alzheimer's disease, Huntington's  
 CC chorea, amyotrophic lateral sclerosis, spinal muscular atrophy and  
 CC Wernicke disease, inflammatory disorders such as nephritis, Crohn's  
 CC disease, ischemia-reperfusion injury, shock, sepsis and inflammatory  
 CC bowel disease. The sequences exhibit activity relating to angiogenesis,  
 CC cell proliferation, cell differentiation, stem cell growth factor,  
 CC activin or inhibin. Therefore, they can be used to manipulate stem cells  
 CC in culture to give rise to neuroepithelial cells that can be used to  
 CC augment or replace cells damaged by illness, accidental damage or genetic  
 CC disorders. The sequences may also be used for regeneration of bone,  
 CC cartilage, tendons and ligaments and in tissue repair and burn healing.  
 CC Note: Some sequences for this patent did not form part of the printed  
 CC specification, but were obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 415 AA;  
 Query Match 100.0%; Score 36; DB 4; Length 415;  
 Best Local Similarity 100.0%; Pred. No. 7.8e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 AGGPCV 6  
 Db 288 AGGPCV 293  
 RESULT 5  
 ADJ58205  
 ID ADJ58205 standard; protein; 415 AA.  
 XX  
 XX ADJ58205;  
 AC  
 XX  
 DT 06-MAY-2004 (first entry)  
 XX  
 DE Human Cezanne-2 C-terminal domain 455-858 SEQ ID NO:5.  
 XX  
 XX enzymatic deubiquitination; TRAFB domain;  
 KW tumor necrosis factor receptor associated factor binding domain;  
 KW Cezanne; core catalytic domain; human; Cezanne-2.  
 XX  
 OS Homo sapiens.  
 OS  
 PN WO2004011636-A2.  
 XX  
 XX 05-FEB-2004.  
 PD  
 XX 31-JUL-2003; 2003WO-GB003320.  
 PF  
 XX 31-JUL-2002; 2002US-0399439P.  
 PR 25-MAR-2003; 2003GB-0006643.

XX  
 PA (BABR-) BABRAHAM INST.  
 XX  
 XX Evans PC, Kilshaw PJ;  
 PI  
 XX WPI; 2004-143859/14.  
 DR  
 XX  
 XX Enzymatic deubiquitination of a molecule which has been ubiquitinated to  
 PT form a ubiquitinated molecule, for stabilising a target molecule,  
 PT comprises contacting the ubiquitinated molecule with an enzyme homologous  
 PT to Cezanne.  
 XX  
 PS Claim 22; SEQ ID NO 5; 80bp; English.  
 CC  
 XX The present invention describes a method for enzymatic deubiquitination  
 CC of a molecule which has been ubiquitinated to form a ubiquitinated  
 CC molecule. The method comprises contacting the molecule with an enzyme or  
 CC domain with 20 % amino acid sequence identity with the TRAFB (tumor  
 CC necrosis factor (TNF) receptor associated factor binding) domain defined  
 CC by Cezanne amino acids 160-416 (ADJ58204), and/or the Cezanne core catalytic  
 CC Cezanne amino acids 126-455 (ADJ58204), and/or the Cezanne core catalytic  
 CC domain defined by Cezanne amino acids 182-455 (ADJ58206). Also described:  
 CC (1) a method for regulating the method described above by contacting a  
 CC molecule which has been ubiquitinated or a deubiquitination enzyme with a  
 CC regulatory polypeptide having at least 20% amino acid sequence identity  
 CC with the C-terminal domain of Cezanne defined by Cezanne amino acid  
 CC residues 444-858 (ADJ58205) or its functional fragment, homologue,  
 CC variant or mutant (for example, a conservatively substituted mutant); (2)  
 CC a method for modulating the activity of a molecule which is capable of  
 CC being ubiquitinated to form a ubiquitinated molecule, comprising  
 CC contacting the molecule or the ubiquitinated molecule with a polypeptide  
 CC cited above, where the polypeptide has deubiquitination activity; (3) a  
 CC protein having an amino acid sequence of Cezanne-2, or its functional  
 CC fragment, homologue, variant or mutant (for example, a conservatively  
 CC substituted mutant); and (4) an isolated nucleotide encoding the protein.  
 CC The methods are useful for stabilising the molecule or target molecule,  
 CC for targeting or maintaining the molecule or target molecule to or at a  
 CC sub-cellular location, and for modulating the activity of the molecule or  
 CC target molecule. The present sequence represents the human Cezanne-2 C-  
 CC terminal domain, which is used in the exemplification of the present  
 CC invention.  
 XX  
 SQ Sequence 415 AA;  
 Query Match 100.0%; Score 36; DB 8; Length 415;  
 Best Local Similarity 100.0%; Pred. No. 7.8e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 AGGPCV 6  
 Db 288 AGGPCV 293  
 RESULT 6  
 ABB97204  
 ID ABB97204 standard; protein; 479 AA.  
 XX  
 XX ABB97204;  
 AC  
 XX  
 DT 28-JUN-2002 (first entry)  
 XX  
 DE Novel human protein SEQ ID NO: 472.  
 XX  
 XX Human; anti-inflammatory; vulnery; anti-inflammatory; immunomodulator;  
 KW anti-infectivity; cerebroprotective; cytoprotective; rheumatic; gene therapy;  
 KW neuroprotective; antiparkinsonian; protein therapy; BBT;  
 KW expressed sequence tag.  
 XX  
 OS Homo sapiens.  
 OS  
 PN WO200222660-A2.  
 XX  
 XX 21-MAR-2002.



XX 10-SEP-2001; 2001MO-US026015.  
PF  
XX 11-SEP-2000; 2000US-00659671.  
PR  
XX (HYSE-) HYSEQ INC.  
PA  
PI Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;  
PI Xue AJ, Yang Y, Wehrman T, Drmanac RT;  
XX WPI; 2002-292408/33.  
XX N-PSDB; ABN32390.  
DR  
XX An isolated polynucleotide for treating diseases associated with its  
PT encoded polypeptide such as cancer and multiple sclerosis.  
XX  
XX Example 2; SEQ ID NO 472; 509bp; English.  
XX  
XX The present invention provides the protein and coding sequences of 444  
CC novel human proteins. These were isolated from expressed sequences tags  
CC (ESTs). They can be used to stimulate cell growth, to regulate  
CC haematopoiesis e.g. to treat aplastic anaemia, to help tissue regrowth  
CC e.g. in burn treatment, to regulate the immune system e.g. to treat  
CC multiple sclerosis, to regulate actiyin or inhibin e.g. to treat  
CC infertility, to regulate haemostasis or thrombolysis e.g. to treat stroke  
CC and cancer, to screen for drugs, to treat inflammatory conditions e.g.  
CC rheumatoid arthritis, and to treat nervous system disorders e.g.  
CC Parkinson's disease. The present sequence is a protein of the invention  
XX  
SQ Sequence 479 AA;  
Query Match Best Local Similarity 100.0%; Score 36; DB 5; Length 479;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 AGGPCV 6  
Db 352 AGGPCV 357  
RESULT 7  
AD126089  
ID AD126089 standard; protein; 843 AA.  
XX  
AC AD126089;  
XX  
XX 22-APR-2004 (first entry)  
DT  
XX  
DE Human protein that promotes STAT6 activation #27.  
XX  
XX human; signal transducer and activator of transcription 6; STAT6;  
KW immunogen; STAT6 activation; allergy; inflammation; autoimmune disease;  
KW diabetes; hyperlipidaemia; infection; cancer; Th1 hyperactive disease;  
KW rheumatoid arthritis; osteoarthritis; systemic lupus erythematosus;  
KW sepsis; asthma; allergic rhinitis; ischemic heart disease;  
KW subarachnoid haemorrhage; viral hepatitis; AIDS.  
XX  
XX Homo sapiens.  
OS  
XX WO2003104277-A2.  
PN  
XX 18-DEC-2003.  
PD  
XX  
XX 05-JUN-2003; 2003WO-JP007123.  
PF  
XX 05-JUN-2002; 2002JP-00164257.  
PR 06-JUN-2002; 2002US-0385912P.  
PR 26-DEC-2002; 2002JP-00377326.  
PR 27-DEC-2002; 2002US-0436467P.  
PR 15-MAY-2003; 2003JP-00137505.  
PR 16-MAY-2003; 2003US-0470836P.  
XX  
XX (ASAH) ASAH KASEI KK.  
PA

XX Sugahara T, Matsuda A, Honda G, Muramatsu S, Ishizawa K;  
PI WPI; 2004-122214/12.  
XX  
XX N-PSDB; AD126088.  
DR  
XX  
XX New signal transducer and activator of transcription 6 activation  
PT promoting purified protein, for diagnosing and treating disease  
PT associated with activation/inhibition of transcription factor e.g.  
PI diabetes and cancer.  
XX  
XX Claim 1; SEQ ID NO 54; 1368bp; English.  
XX  
XX The invention relates to a purified protein promoting signal transducer  
CC and activator of transcription 6 activation (STAT6). The protein is  
CC useful for the producing an antibody, which involves administering the  
CC protein or its epitope-bearing fragments to a non-human animal as an  
CC antigen. The nucleic acid is useful for diagnosing a disease or  
CC susceptibility to a disease related to expression or activity of the  
CC protein. A transformant expressing the protein is useful for screening  
CC compounds which inhibit or promote STAT6 activation. A transformant  
CC expressing the protein is useful for producing a pharmaceutical  
CC composition. Compositions, antibodies and antisense molecules are useful  
CC for the treating a disease associated with STAT6 activation such as  
CC allergic diseases, inflammation, autoimmune diseases, diabetes,  
CC hyperlipidaemia, infectious disease and cancers. Compositions are useful  
CC for treating diseases associated with STAT6 activation and/or prevention  
CC of Th1 hyperactive diseases. Compositions are also useful in rheumatoid  
CC arthritis, osteoarthritis, systemic lupus erythematosus, sepsis, asthma,  
CC allergic rhinitis, ischemic heart disease, subarachnoid haemorrhage,  
CC viral hepatitis and AIDS. The protein has efficient promoting STAT6  
CC activity. The protein or nucleic acid is effectively useful for screening  
CC compounds for treating and preventing disease associated with excessive  
CC activation or inhibition of STAT6. The present sequence represents the  
CC amino acid sequence of a human protein which promotes STAT6 activation.  
XX  
SQ Sequence 843 AA;  
Query Match Best Local Similarity 100.0%; Score 36; DB 8; Length 843;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 AGGPCV 6  
Db 716 AGGPCV 721  
RESULT 8  
AD126087  
ID AD126087 standard; protein; 843 AA.  
XX  
XX AD126087;  
AC  
XX  
XX 22-APR-2004 (first entry)  
DT  
XX  
DE Human protein that promotes STAT6 activation #26.  
XX  
XX human; signal transducer and activator of transcription 6; STAT6;  
KW immunogen; STAT6 activation; allergy; inflammation; autoimmune disease;  
KW diabetes; hyperlipidaemia; infection; cancer; Th1 hyperactive disease;  
KW rheumatoid arthritis; osteoarthritis; systemic lupus erythematosus;  
KW sepsis; asthma; allergic rhinitis; ischemic heart disease;  
KW subarachnoid haemorrhage; viral hepatitis; AIDS.  
XX  
XX Homo sapiens.  
OS  
XX WO2003104277-A2.  
PN  
XX 18-DEC-2003.  
PD  
XX  
XX 05-JUN-2003; 2003WO-JP007123.  
PF  
XX 05-JUN-2002; 2002JP-00164257.  
PR

PR 06-JUN-2002; 2002US-0385912P.  
 PR 26-DEC-2002; 2002JP-00377326.  
 PR 27-DEC-2002; 2002US-0436467P.  
 PR 15-MAY-2003; 2003JP-00137505.  
 PR 16-MAY-2003; 2003US-0470836P.  
 PA (ASAH) ASAMI KASEI KK.  
 PI Sugahara T, Matsuda A, Honda G, Muramatsu S, Ishizawa K;  
 DR WPI; 2004-122214/12.  
 DR N-PSDB; ADI26086.  
 XX  
 PT New signal transducer and activator of transcription 6 activation  
 PT promoting purified protein, for diagnosing and treating disease  
 PT associated with activation/inhibition of transcription factor e.g.  
 PT diabetes and cancer.  
 PS Claim 1; SEQ ID NO 52; 1368pp; English.  
 XX  
 CC The invention relates to a purified protein promoting signal transducer  
 CC and activator of transcription 6 activation (STAT6). The protein is  
 CC useful for the producing an antibody, which involves administering the  
 CC protein or its epitope-bearing fragments to a non-human animal as an  
 CC antigen. The nucleic acid is useful for diagnosing a disease or  
 CC susceptibility to a disease related to expression or activity of the  
 CC protein. A transformant expressing the protein is useful for screening  
 CC compounds which inhibit or promote STAT6 activation. A transformant  
 CC expressing the protein is useful for producing a pharmaceutical  
 CC composition. Compositions, antibodies and antisense molecules are useful  
 CC for the treating a disease associated with STAT6 activation such as  
 CC allergic diseases, inflammation, autoimmune diseases, diabetes,  
 CC hyperlipidemia, infectious diseases and cancers. Compositions are useful  
 CC for treating disease associated with STAT6 activation and/or prevention  
 CC of Th1 hyperactive diseases. Compositions are also useful in rheumatoid  
 CC arthritis, osteoarthritis, systemic lupus erythematosus, sepsis, asthma,  
 CC allergic rhinitis, ischemic heart diseases, subarachnoid hemorrhage,  
 CC viral hepatitis and AIDS. The protein has efficiently promoting STAT6  
 CC activity. The protein or nucleic acid is effectively useful for screening  
 CC compounds for treating and preventing disease associated with excessive  
 CC activation or inhibition of STAT6. The present sequence represents the  
 CC amino acid sequence of a human protein which promotes STAT6 activation.  
 XX  
 SQ Sequence 843 AA;  
 Query Match 100.0%; Score 36; DB 8; Length 843;  
 Best Local Similarity 100.0%; Pred. No. 1.4e+03;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 AGPCV 6  
 Db 716 AGPCV 721  
 RESULT 9  
 ABR47408  
 ID ABR47408 standard; protein; 858 AA.  
 XX  
 AC ABR47408;  
 DT 12-JUN-2003 (first entry)  
 XX  
 DE Breast cancer associated protein sequence SEQ ID NO:47.  
 XX  
 KM Human; breast cancer; cytostatic; gene therapy.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2003004989-A2.  
 XX  
 PD 16-JAN-2003.  
 XX  
 PF 21-JUN-2002; 2002WO-US019669.

XX 21-JUN-2001; 2001US-0299887P.  
 PR 27-JUN-2001; 2001US-0301572P.  
 PR 18-JUL-2001; 2001US-0306501P.  
 PR 25-SEP-2001; 2001US-0325002P.  
 PR 05-MAR-2002; 2002US-0362585P.  
 PR 14-MAY-2002; 2002US-0380391P.  
 XX  
 PA (MILL-) MILLENIUM PHARM INC.  
 PI Lillie J, Gannavarapu M, Glatt K, Hoersch S, Kamatkar S;  
 PI Mertens M, Monahan JE, Myer V, Wang Y, Xu Y, Zhao X, Meyers RB,  
 PI Bast RC, Horrobagyl GN, Puzsai L, Meric F, Sahin A, Mills GB;  
 XX  
 DR WPI; 2003-210381/20.  
 DR N-PSDB; ACC50099.  
 XX  
 PT Breast cancer diagnosis or treatment by comparing the level of expression  
 PT of a marker in a patient sample with that in the control non-breast  
 PT cancer sample.  
 PS Claim 1; SEQ ID NO 47; 128pp; English.  
 XX  
 CC The present invention describes a method for assessing whether a patient  
 CC is afflicted with breast cancer. The method comprises comparing the level  
 CC of expression of a marker (gene/polypeptide see ACC50076 to ACC50334 and  
 CC ABR4736 to ABR47632) in a patient sample and the normal level of  
 CC expression of the marker in a control non-breast cancer sample, where a  
 CC significant increase in the level of expression of the marker in the  
 CC patient sample and the normal level is an indication that the patient is  
 CC afflicted with breast cancer. The breast cancer associated sequences from  
 CC the present invention have cytostatic activities and can be used in gene  
 CC therapy. The method is useful for diagnosing and treating breast cancer.  
 CC N.B. The sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 858 AA;  
 Query Match 100.0%; Score 36; DB 6; Length 858;  
 Best Local Similarity 100.0%; Pred. No. 1.5e+03;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 AGPCV 6  
 Db 731 AGPCV 736  
 RESULT 10  
 ADC37584  
 ID ADC37584 standard; protein; 858 AA.  
 XX  
 AC ADC37584;  
 DT 18-DEC-2003 (first entry)  
 XX  
 DE Human nucleic acid associated protein, NAAp-51.  
 XX  
 KM Human; nucleic acid associated protein; NAAp; cytostatic;  
 KM antiarteriosclerotic; anticonvulsant; neurotropic; neuroprotective;  
 KM cerebroprotective; anti-HIV; antiallergic; antiinflammatory;  
 KM thyromimetic; gene therapy; cell proliferative disorder; cancer;  
 KM atherosclerosis; neurological disorder; epilepsy; Huntington's disease;  
 KM stroke; immune disorder; inflammatory disorder; AIDS; allergy;  
 KM developmental disorder; Hypothyroidism; Cushing's syndrome; infection.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2003046151-A2.  
 XX  
 PD 05-JUN-2003.  
 XX  
 PF 26-NOV-2002; 2002WO-US038445.

XX 27-NOV-2001; 2001US-0333925P.  
 PR 07-DEC-2001; 2001US-0340477P.  
 PR 14-DEC-2001; 2001US-0340362P.  
 PR 18-DEC-2001; 2001US-0342002P.  
 PA (INCYTE GENOMICS INC.)  
 XX  
 PI Baughn MR, Becha SD, Bhattacharya U, Blake JU, Burford N, Burrill JD,  
 PI Chang H, Chawla NK, Elliott VS, Emerling BM, Forgyte TJ, Gaudin AR,  
 PI Gleason KJ, Gorvay AE, Griffin JA, Hatfield AJ, Jackson JL, Ho A,  
 PI Ison GH, Jackson AA, Jiang X, Jin P, Kabile AE, Khare R, Lal PG,  
 PI Lee S, Lee SY, Li JX, Lu DM, Ramkumar J, Richardson TW,  
 PI Sprague WW, Swarnakar A, Tang YT, Warren BA, Xu Y, Yao MG, Yue H,  
 PI Zheng W;  
 XX WPI; 2003-513642/48.  
 DR N-PSDB; ADC37644.  
 XX  
 PT New human nucleic acid associated proteins (NAAP), useful for diagnosing,  
 PT treating and preventing diseases or conditions associated with the  
 PT aberrant NAAP expression e.g. cancer, AIDS, atherosclerosis, epilepsy, or  
 PT infections.  
 PS Claim 1; SEQ ID NO 51; 383bp; English.  
 XX  
 CC The present invention relates to novel human nucleic acid associated  
 CC protein (NAAP) (1, ADC37534-ADC37593) and their coding sequences (1). The  
 CC NAAPs and their coding sequences are useful in diagnosing, treating and  
 CC preventing diseases or conditions associated with the decreased  
 CC expression or over expression of NAAP, such as cell proliferative (e.g.  
 CC cancer, atherosclerosis), neurological (e.g. epilepsy, Huntington's  
 CC disease, stroke), immune/inflammatory (e.g. AIDS, allergies) and  
 CC developmental (e.g. Hypothyroidism, Cushing's syndrome) disorders, or  
 CC infections. These are also useful in assessing the effects of exogenous  
 CC compounds on the expression of nucleic acid and amino acid sequences of  
 CC NAAP.  
 CC  
 XX  
 SQ Sequence 858 AA;  
 XX  
 Query Match 100.0%; Score 36; DB 7; Length 858;  
 Best Local Similarity 100.0%; Pred. No. 1.5e+03;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 AGGPGCV 6  
 Db 731 AGGPGCV 736  
 XX  
 RESULT 11  
 ADF76869  
 ID ADF76869 standard; protein; 1745 AA.  
 XX  
 AC ADF76869;  
 XX  
 DT 26-FEB-2004 (first entry)  
 XX  
 DE Novel human secreted and transmembrane protein SeqID 544.  
 XX  
 KW human; PRO; membrane bound protein; membrane bound receptor;  
 KW cell proliferation; cell migration; cell differentiation;  
 KW mitogenic factor; survival factor; cytotoxic factor;  
 KW differentiation factor; neuropeptide; hormone; cell receptor;  
 KW receptor-ligand interaction; cytoskeletal; chondrocyte; tumour.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2003072035-A2.  
 XX  
 PD 04-SBP-2003.  
 XX  
 PF 21-FEB-2003; 2003WO-US005241.  
 XX

PR 22-FEB-2002; 2002US-0359461P.  
 XX  
 PA (GERTH) GENENTECH INC.  
 XX  
 PI Bodary SC, Clark H, Hunte B, Jackman JK, Schoenfeld JR,  
 PI Williams PM, Wood WI, Wu TD,  
 PI WPI; 2003-721702/68.  
 DR N-PSDB; ADF76868.  
 XX  
 PT New PRO polypeptides, useful for diagnosing and treating an immune  
 PT related disorder, e.g. systemic lupus erythematosus, rheumatoid  
 PT arthritis, osteoarthritis, juvenile chronic arthritis, thyroiditis or  
 PT diabetes mellitus.  
 PS Claim 10; SEQ ID NO 544; 918bp; English.  
 XX  
 CC This invention relates to novel nucleic acids encoding human PRO secreted  
 CC and transmembrane proteins. Extracellular proteins play important roles  
 CC in the formation, differentiation and maintenance of multicellular  
 CC organisms. The fate of many individual cells (for example proliferation,  
 CC migration or differentiation) is typically governed by information  
 CC received from other cells and the immediate environment. The information  
 CC is often transmitted by secreted polypeptides (for example mitogenic  
 CC factors, survival factors, cytotoxic factors, differentiation factors,  
 CC neuropeptides and hormones) which are received and interpreted by diverse  
 CC cell receptors or membrane bound proteins. These membrane bound proteins  
 CC and receptors may be of use as pharmaceutical and diagnostic agents, such  
 CC as in the blocking of receptor-ligand interactions. The current invention  
 CC provides the amino acid sequences of novel human membrane bound receptors  
 CC and proteins, along with the cDNA sequences encoding them. The novel  
 CC proteins of the invention may have cytostatic activities through the  
 CC stimulation of chondrocytes. The nucleic acids of the invention may be  
 CC useful for the manufacture of a medicament for diagnosing or treating a  
 CC tumour in a mammal. In addition, they may be useful for measuring or  
 CC detecting the expression of a tumour associated gene. The present  
 CC sequence is the amino acid sequence of a human PRO protein of the  
 CC invention.  
 CC  
 XX  
 SQ Sequence 1745 AA;  
 XX  
 Query Match 100.0%; Score 36; DB 7; Length 1745;  
 Best Local Similarity 100.0%; Pred. No. 2.7e+03;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 AGGPGCV 6  
 Db 1 AGGPGCV 6  
 XX  
 RESULT 12  
 ADO71781  
 ID ADO71781 standard; protein; 1745 AA.  
 XX  
 AC ADO71781;  
 XX  
 DT 26-AUG-2004 (first entry)  
 XX  
 DE BFLP0169 related protein, SEQ ID 21.  
 XX  
 KW Antiinflammatory; Dermatologically; Immunosuppressive; human; BFLP0169;  
 KW kidney; lupus nephritis; autoimmune disorder; lupus.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2004048521-A2.  
 XX  
 PD 10-JUN-2004.  
 XX  
 PF 21-NOV-2003; 2003WO-US037339.  
 XX  
 PR 21-NOV-2002; 2002US-0428094P.  
 XX

PA (AMHP) WYETH.  
 XX O'loole M, Mounts WM, Shojasee N;  
 XX  
 DR WPI; 2004-441163/41.  
 XX  
 PT New isolated BFLP0169 nucleic acids and proteins, useful for diagnosing  
 XX or treating autoimmune disorder, e.g. lupus or lupus nephritis.  
 XX  
 PS Disclosure; Page 9-13; 83pp; English.  
 XX  
 CC The present invention relates to human BFLP0169 (ADO71763) and its coding  
 CC sequence (ADO71762). BFLP0169 gene expression is increased in kidney  
 CC tissue in mice with lupus nephritis. The BFLP0169 sequences are useful  
 CC for diagnosing and treating autoimmune disorder, preferably lupus or  
 CC lupus nephritis. They are also useful in the generation of antibodies  
 CC that bind immunospecifically to the nucleic acids. The present sequence  
 CC was used in a sequence alignment with the BFLP0169 protein.  
 XX  
 SQ Sequence 1745 AA;  
 Query Match 100.0%; Score 36; DB 8; Length 1745;  
 Best Local Similarity 100.0%; Pred. No. 2.7e+03;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 AGGPGCV 6  
 DB 1 AGGPGCV 6  
 RESULT 13  
 ID ADO71782 standard; protein; 1767 AA.  
 XX ADO71782;  
 AC  
 XX 26-AUG-2004 (first entry)  
 DT  
 XX BFLP0169 related protein, SEQ ID 22.  
 DE  
 XX  
 KM Antinflammatory; Dermatological; Immunosuppressive; human; BFLP0169;  
 XX kidney; lupus nephritis; autoimmune disorder; lupus.  
 OS  
 XX Homo sapiens.  
 XX  
 PN WO2004048521-A2.  
 XX  
 PD 10-JUN-2004.  
 XX  
 PF 21-NOV-2003; 2003WO-US037339.  
 XX  
 PR 21-NOV-2002; 2002US-0428094P.  
 XX  
 PA (AMHP) WYETH.  
 XX  
 PI O'loole M, Mounts WM, Shojasee N;  
 XX  
 DR WPI; 2004-441163/41.  
 XX  
 PT New isolated BFLP0169 nucleic acids and proteins, useful for diagnosing  
 XX or treating autoimmune disorder, e.g. lupus or lupus nephritis.  
 XX  
 PS Disclosure; Page 9-13; 83pp; English.  
 XX  
 CC The present invention relates to human BFLP0169 (ADO71763) and its coding  
 CC sequence (ADO71762). BFLP0169 gene expression is increased in kidney  
 CC tissue in mice with lupus nephritis. The BFLP0169 sequences are useful  
 CC for diagnosing and treating autoimmune disorder, preferably lupus or  
 CC lupus nephritis. They are also useful in the generation of antibodies  
 CC that bind immunospecifically to the nucleic acids. The present sequence  
 CC was used in a sequence alignment with the BFLP0169 protein.  
 XX  
 SQ Sequence 1767 AA;

Query Match 100.0%; Score 36; DB 8; Length 1767;  
 Best Local Similarity 100.0%; Pred. No. 2.7e+03;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 AGGPGCV 6  
 DB 19 AGGPGCV 24  
 RESULT 14  
 ID ABG09906 standard; protein; 1889 AA.  
 XX ABG09906;  
 AC  
 XX 13-FEB-2002 (first entry)  
 DT  
 XX  
 DE Novel human diagnostic protein #9897.  
 XX  
 KM Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 XX food supplement; medical imaging; diagnostic; genetic disorder.  
 OS  
 XX Homo sapiens.  
 XX  
 PN WO200175067-A2.  
 XX  
 PD 11-OCT-2001.  
 XX  
 PF 30-MAR-2001; 2001WO-US008631.  
 XX  
 PR 31-MAR-2000; 2000US-00540217.  
 XX  
 PR 23-AUG-2000; 2000US-00649167.  
 XX  
 PA (HSEB-) HSEB INC.  
 XX  
 PI Drmanac RT, Liu C, Tang YT;  
 XX  
 DR WPI; 2001-639362/73.  
 XX  
 DR N-PSDB; AAS74093.  
 XX  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 XX diagnostics, forensics, gene mapping, identification of mutations  
 XX responsible for genetic disorders or other traits and to assess  
 XX biodiversity.  
 PT  
 PS Claim 20; SEQ ID NO 40265; 103pp; English.  
 XX  
 CC The invention relates to isolated polynucleotide (I) and polypeptide (II)  
 CC sequences. (I) is useful as hybridisation probes, polymerase chain  
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
 CC and in recombinant production of (II). The polynucleotides are also used  
 CC in diagnostics as expressed sequence tags for identifying expressed  
 CC genes. (I) is useful in gene therapy techniques to restore normal  
 CC activity of (II) or to treat disease states involving (II). (II) is  
 CC useful for generating antibodies against it, detecting or quantitating a  
 CC polypeptide in tissue, as molecular weight markers and as a food  
 CC supplement. (II) and its binding partners are useful in medical imaging  
 CC of sites expressing (II). (I) and (II) are useful for treating disorders  
 CC involving aberrant protein expression or biological activity. The  
 CC polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG03037 represent novel human diagnostic  
 CC amino acid sequences of the invention. Note: The sequence data for this  
 CC patent did not appear in the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 1889 AA;  
 Query Match 100.0%; Score 36; DB 4; Length 1889;

Best Local Similarity 100.0%; Pred. No. 2.9e+03;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGPCV 6  
DB 1 AGPCV 6

## RESULT 15

AA0795569  
ID AAG795569 standard; peptide, 142 AA.

AC AAG795569;

DT 23-DEC-2002 (first entry)

DE ACK1 C-terminal serine/proline-rich region.

XX Mlg-6; lung; fibroblast; gene 33; rat; hyperproliferation;

KM serine/threonine phosphorylation site; inflammation; tumor; Erk2;

KM Cdc42 and rac interaction binding; CRIB; SH3; WW; binding motif;

KM tyrosine kinase; ACK1; epidermal growth factor receptor; EGFR;

XX mitogen-activated protein kinase pathway.

OS Rattus rattus.

PN EPI236474-A1.

XX 04-SEP-2002.

PD 26-FEB-2001; 2001EP-00104737.

XX 26-FEB-2001; 2001EP-00104737.

PR 26-FEB-2001; 2001EP-00104737.

XX (PIAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.

PA Ullrich A, Hackel P;

XX WPI; 2002-692997/75.

DR WPI; 2002-692997/75.

XX WPI; 2002-692997/75.

XX WPI; 2002-692997/75.

XX WPI; 2002-692997/75.

XX WPI; 2002-692997/75.

XX WPI; 2002-692997/75.

XX WPI; 2002-692997/75.

XX WPI; 2002-692997/75.

XX WPI; 2002-692997/75.

XX WPI; 2002-692997/75.

XX WPI; 2002-692997/75.

XX WPI; 2002-692997/75.

XX WPI; 2002-692997/75.

XX WPI; 2002-692997/75.

XX WPI; 2002-692997/75.

XX WPI; 2002-692997/75.

RESULT 16  
ABR82256  
ID ABR82256 standard; protein, 1036 AA.

XX ABR82256;

DT 13-OCT-2003 (first entry)

DE Human activated p12cdc42hs kinase (ACK1).

XX ACK1; activated p12cdc42hs kinase; cytosolic; gene therapy; human;

KM immune response inducer; cancer; enzyme.

XX Homo sapiens.

OS WO2003054512-A2.

PN 03-JUL-2003.

XX 13-DEC-2002; 2002MO-US039927.

XX 20-DEC-2001; 2001US-0341436P.

XX (TULIA-) TULARIK INC.

XX Degenhardt YV, Powers S;

XX WPI; 2003-569275/53.

XX N-PSDB; ACC85084.

XX WPI; 2003-569275/53.

XX WPI; 2003-569275/53.

XX WPI; 2003-569275/53.

XX WPI; 2003-569275/53.

XX WPI; 2003-569275/53.

XX WPI; 2003-569275/53.

XX WPI; 2003-569275/53.

XX WPI; 2003-569275/53.

XX WPI; 2003-569275/53.

XX WPI; 2003-569275/53.

XX WPI; 2003-569275/53.

XX WPI; 2003-569275/53.

XX WPI; 2003-569275/53.

XX WPI; 2003-569275/53.

XX WPI; 2003-569275/53.

XX WPI; 2003-569275/53.

XX WPI; 2003-569275/53.

XX WPI; 2003-569275/53.

XX WPI; 2003-569275/53.

XX WPI; 2003-569275/53.

XX WPI; 2003-569275/53.

XX WPI; 2003-569275/53.

XX WPI; 2003-569275/53.

RESULT 17  
ADE08157  
ID ADE08157 standard; protein, 1068 AA.

XX ADE08157;

DT 29-JAN-2004 (first entry)

DE Human activated p12cdc42hs kinase (ACK1).

XX ACK1; activated p12cdc42hs kinase; cytosolic; gene therapy; human;

KM immune response inducer; cancer; enzyme.

XX Homo sapiens.

OS WO2003054512-A2.

PN 03-JUL-2003.

XX 13-DEC-2002; 2002MO-US039927.

XX 20-DEC-2001; 2001US-0341436P.

XX (TULIA-) TULARIK INC.

XX Degenhardt YV, Powers S;

XX WPI; 2003-569275/53.

XX N-PSDB; ACC85084.

XX WPI; 2003-569275/53.

XX WPI; 2003-569275/53.

XX WPI; 2003-569275/53.

XX WPI; 2003-569275/53.

XX WPI; 2003-569275/53.

XX WPI; 2003-569275/53.

XX WPI; 2003-569275/53.

XX WPI; 2003-569275/53.

XX WPI; 2003-569275/53.

XX WPI; 2003-569275/53.

XX WPI; 2003-569275/53.

XX WPI; 2003-569275/53.

XX WPI; 2003-569275/53.

XX WPI; 2003-569275/53.

XX WPI; 2003-569275/53.

XX WPI; 2003-569275/53.

XX WPI; 2003-569275/53.

XX WPI; 2003-569275/53.

XX WPI; 2003-569275/53.

XX WPI; 2003-569275/53.

XX WPI; 2003-569275/53.

XX WPI; 2003-569275/53.



PT New peptide compounds useful for detecting herpes simplex virus type 2 -  
 PT can differentiate between herpes simplex virus type 1 and type 2.  
 XX  
 PS Example 1, Page 10; 25pp; English.  
 XX  
 CC The invention relates to a multiply displayed peptide structure of  
 CC formula [(X1)p-A-(X2)q-SP]n-Core, X1, X2 = 1-6 non-interfering amino acid  
 CC residues; A = Glu Gly Phe Glu Gly Ala Gly Asp Gly Pro Gly Asp Asp  
 CC Asp; Sp = spacer group extending outwardly from the core; n at least 4;  
 CC and p, q = 0 or 1. The linkage between the core and the spacer may be  
 CC chemical or physical. Peptides of this formula are used in the diagnosis  
 CC of herpes simplex virus type 2 (HSV-2). The peptides can distinguish HSV-  
 CC 2 from HSV-1. A series of 67 peptides (AAW6624-W6690), mostly 18 amino  
 CC acids long, that spanned amino acids 21-69 of the predicted open reading  
 CC frame of HSV gG2 were synthesized. The peptides were made as multiply  
 CC displayed peptide structures of the invention and were screened against  
 CC sera from HSV-1, HSV-2 antibody-positive individuals and from individuals  
 CC having no laboratory evidence of HSV infection. From the results, peptide  
 CC 55 (AAW6678) was considered a likely candidate for type specific  
 CC serodiagnosis of HSV  
 XX  
 SQ Sequence 18 AA;  
 Query Match 88.9%; Score 32; DB 2; Length 18;  
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 GGPCV 6  
 |||||  
 DB 5 GGPCV 9  
 RESULT 20  
 AAR45552  
 ID AAR45552 standard; protein, 20 AA.  
 XX  
 AC AAR45552;  
 XX  
 DT 25-MAR-2003 (revised)  
 DT 13-JUL-1994 (first entry)  
 XX  
 DE Cry j I pollen allergen peptide CJI-11.  
 XX  
 KW Japanese cedar; detection; allergy; treatment; diagnosis; T cell epitope;  
 KW sensitivity.  
 XX  
 OS Cryptomeria japonica.  
 XX  
 PN WO9401560-A1.  
 XX  
 PD 20-JAN-1994.  
 XX  
 PF 15-JAN-1993; 93WO-US000139.  
 XX  
 PR 10-JUL-1992; 92WO-US005661.  
 PR 01-SEP-1992; 92US-00938990.  
 XX  
 PA (IMMU-) IMMUNOLOGIC PHARM CORP.  
 XX  
 PI Griffith J, Pollock J, Bond JF, Garman RD, Kuo M;  
 DR WPI, 1994-035066/04.  
 XX  
 PT Antigens derived from Japanese cedar pollen allergen Cry j I - contain at  
 PT least two T cell epitope(s), used to treat or diagnose allergy.  
 XX  
 PS Claim 1; Fig 13; 137pp; English.  
 XX  
 CC The sequence is that of an isolated peptide of the Japanese cedar pollen  
 CC allergen Cry j I (amino acids 101-120). The peptide, CJI-11, can be used  
 CC for the treatment and diagnosis of allergies associated with Japanese  
 CC cedar pollen. It has enhanced therapeutic properties but reduced side  
 CC effects compared to naturally occurring allergens. (Updated on 25-MAR-

CC 2003 to correct PN field.)  
 XX  
 SQ Sequence 20 AA;  
 Query Match 88.9%; Score 32; DB 2; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 2.5e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 GGPCV 6  
 |||||  
 DB 4 GGPCV 8  
 RESULT 21  
 AAR45551  
 ID AAR45551 standard; protein, 20 AA.  
 XX  
 AC AAR45551;  
 XX  
 DT 25-MAR-2003 (revised)  
 DT 13-JUL-1994 (first entry)  
 XX  
 DE Cry j I pollen allergen peptide CJI-10.  
 XX  
 KW Japanese cedar; detection; allergy; treatment; diagnosis; T cell epitope;  
 KW sensitivity.  
 XX  
 OS Cryptomeria japonica.  
 XX  
 PN WO9401560-A1.  
 XX  
 PD 20-JAN-1994.  
 XX  
 PF 15-JAN-1993; 93WO-US000139.  
 XX  
 PR 10-JUL-1992; 92WO-US005661.  
 PR 01-SEP-1992; 92US-00938990.  
 XX  
 PA (IMMU-) IMMUNOLOGIC PHARM CORP.  
 XX  
 PI Griffith J, Pollock J, Bond JF, Garman RD, Kuo M;  
 DR WPI, 1994-035066/04.  
 XX  
 PT Antigens derived from Japanese cedar pollen allergen Cry j I - contain at  
 PT least two T cell epitope(s), used to treat or diagnose allergy.  
 XX  
 PS Claim 1; Fig 13; 137pp; English.  
 XX  
 CC The sequence is that of an isolated peptide of the Japanese cedar pollen  
 CC allergen Cry j I (amino acids 91-110). The peptide, CJI-10, can be used  
 CC for the treatment and diagnosis of allergies associated with Japanese  
 CC cedar pollen. It has enhanced therapeutic properties but reduced side  
 CC effects compared to naturally occurring allergens. (Updated on 25-MAR-  
 CC 2003 to correct PN field.)  
 XX  
 SQ Sequence 20 AA;  
 Query Match 88.9%; Score 32; DB 2; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 2.5e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 GGPCV 6  
 |||||  
 DB 14 GGPCV 18  
 RESULT 22  
 AAR82501  
 ID AAR82501 standard; protein, 20 AA.  
 XX  
 AC AAR82501;  
 XX

DT 27-AUG-2003 (revised)  
 DT 15-APR-1996 (first entry)  
 XX  
 DE Cry j I Japanese Cedar pollen allergen peptide fragment (CJI-11).  
 XX  
 KM Cry j I; Japanese cedar pollen allergen; modified; drug production;  
 KM allergy; Cryptomeria japonica.  
 XX  
 OS Cryptomeria japonica.  
 XX  
 PN WO9527786-A1.  
 PD 19-OCT-1995.  
 XX  
 PF 06-APR-1995; 95WO-US004249.  
 XX  
 PR 08-APR-1994; 94US-00226248.  
 PR 06-DEC-1994; 94US-00350225.  
 XX  
 PA (IMMU-) IMMUNOLOGIC PHARM CORP.  
 XX  
 PI Franzen HM, Powers SP, Kuo M, Evans S, Shaked Z, Chen X;  
 DR WPI; 1995-366391/47.  
 XX  
 PT Modified Cryptomeria japonica (Cry j) I peptide(s) - useful for treating  
 PT allergy to Japanese cedar pollen allergen or immunologically cross  
 PT reactive allergens.  
 XX  
 PS Disclosure; Fig 2; 60pp; English.  
 XX  
 SQ Novel peptides of cry j I have been modified as a part of a  
 CC preformulation scheme to develop an optimised drug product for  
 CC therapeutic treatment of humans suffering from allergy to Japanese cedar  
 CC pollen allergen or an allergen which is immunologically cross reactive  
 CC with Japanese cedar pollen allergen. Such modified peptides possess  
 CC certain characteristics which render them particularly suitable for drug  
 CC product formulation. Peptide fragments of Cry j I, modified and  
 CC unmodified, are given in AAR82491-R82525. This peptide fragment  
 CC corresponds to amino acids 101-120 of the allergen mature protein.  
 CC (Updated on 27-AUG-2003 to correct OS field.)  
 CC  
 SQ Sequence 20 AA;  
 XX  
 Query Match 88.9%; Score 32; DB 2; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 2.5e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 GGPCV 6  
 DB 4 GGPCV 8  
 XX  
 RESULT 23  
 AAR82500  
 ID AAR82500 standard; protein; 20 AA.  
 XX  
 AC AAR82500;  
 XX  
 DT 27-AUG-2003 (revised)  
 DT 15-APR-1996 (first entry)  
 XX  
 DE Cry j I Japanese Cedar pollen allergen peptide fragment (CJI-10).  
 XX  
 KM Cry j I; Japanese cedar pollen allergen; modified; drug production;  
 KM allergy; Cryptomeria japonica.  
 XX  
 OS Cryptomeria japonica.  
 XX  
 PN WO9527786-A1.  
 PD 19-OCT-1995.  
 XX

PF 06-APR-1995; 95WO-US004249.  
 XX  
 PR 08-APR-1994; 94US-00226248.  
 PR 06-DEC-1994; 94US-00350225.  
 XX  
 PA (IMMU-) IMMUNOLOGIC PHARM CORP.  
 XX  
 PI Franzen HM, Powers SP, Kuo M, Evans S, Shaked Z, Chen X;  
 DR WPI; 1995-366391/47.  
 XX  
 PT Modified Cryptomeria japonica (Cry j) I peptide(s) - useful for treating  
 PT allergy to Japanese cedar pollen allergen or immunologically cross  
 PT reactive allergens.  
 XX  
 PS Disclosure; Fig 2; 60pp; English.  
 XX  
 SQ Novel peptides of cry j I have been modified as a part of a  
 CC preformulation scheme to develop an optimised drug product for  
 CC therapeutic treatment of humans suffering from allergy to Japanese cedar  
 CC pollen allergen or an allergen which is immunologically cross reactive  
 CC with Japanese cedar pollen allergen. Such modified peptides possess  
 CC certain characteristics which render them particularly suitable for drug  
 CC product formulation. Peptide fragments of Cry j I, modified and  
 CC unmodified, are given in AAR82491-R82525. This peptide fragment  
 CC corresponds to amino acids 91-110 of the allergen mature protein.  
 CC (Updated on 27-AUG-2003 to correct OS field.)  
 CC  
 SQ Sequence 20 AA;  
 XX  
 Query Match 88.9%; Score 32; DB 2; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 2.5e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 GGPCV 6  
 DB 14 GGPCV 18  
 XX  
 RESULT 24  
 ADD93845  
 ID ADD93845 standard; peptide; 20 AA.  
 XX  
 AC ADD93845;  
 XX  
 DT 29-JAN-2004 (first entry)  
 XX  
 DE Japanese cedar pollen allergen Cry j I T-cell epitope, SEQ ID NO:11.  
 XX  
 KM Japanese cedar; pollen allergen; Cry j I; T-cell epitope; dog;  
 KM Canis familiaris; allergic dermatitis; veterinary; DNA vaccine;  
 KM dermatological; antiinflammatory; vaccine; gene therapy.  
 XX  
 OS Cryptomeria japonica.  
 XX  
 PN JP2003116556-A.  
 PD 22-APR-2003.  
 XX  
 PF 09-OCT-2001; 2001JP-00311433.  
 PR 09-OCT-2001; 2001JP-00311433.  
 XX  
 PA (NIPZ) NIPPON ZENYAKU KOGYO KK.  
 XX  
 DR WPI; 2003-771273/73.  
 XX  
 PT New cedar pollinosis antigen T-cell epitope in a dog useful for preparing  
 PT a DNA vaccine that can be used to prevent allergic dermatitis.  
 XX  
 PS Claim 1; SEQ ID NO 11; 20pp; Japanese.  
 XX  
 CC The invention relates to T-cell epitopes of the Japanese cedar



CC (Cryptomeria japonica) pollen allergen Cry j 1 which stimulate an immune  
 CC response in dogs. The invention also encompasses DNA encoding the Cry j 1  
 CC T-cell epitopes, a vector comprising a Cry j 1 T-cell epitope-encoding  
 CC DNA, and DNA vaccine containing a vector of the invention for the  
 CC treatment of allergic dermatitis. The Cry j 1 T-cell epitopes and their  
 CC encoding nucleic acids are useful in the preparation of vaccines,  
 CC especially DNA vaccines, for the treatment of allergic dermatitis in  
 CC dogs. The DNA vaccine of the invention can be prepared inexpensively with  
 CC easier manufacturing and purification processes. The present sequence  
 CC represents a specifically claimed Cry j 1 T-cell epitope of the  
 CC invention.

XX  
 XX Sequence 20 AA;

Query Match 88.9%; Score 32; DB 7; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 2.5e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GGPCV 6  
 |||||  
 Db 4 GGPCV 8

RESULT 25

ADD93844 standard; peptide; 20 AA.

AC ADD93844;

DT 29-JAN-2004 (first entry)

DE Japanese cedar pollen allergen Cry j 1 T-cell epitope. SEQ ID NO:10.

KW Japanese cedar; pollen allergen; Cry j 1; T-cell epitope; dog;  
 KM Canis familiaris; allergic dermatitis; veterinary; DNA vaccine;  
 XX dermatological; antiinflammatory; vaccine; gene therapy.

OS Cryptomeria japonica.

PN JP2003116556-A.

PD 22-APR-2003.

PF 09-OCT-2001; 2001JP-00311433.

PR 09-OCT-2001; 2001JP-00311433.

PA (NIPZ ) NIPPON ZENYAKU KOGYO KK.

DR WPI; 2003-771273/73.

PT New cedar pollinosis antigen T-cell epitope in a dog useful for preparing  
 PT a DNA vaccine that can be used to prevent allergic dermatitis.

PS Claim 1; SEQ ID NO 10; 20pp; Japanese.

XX The invention relates to T-cell epitopes of the Japanese cedar  
 CC (Cryptomeria japonica) pollen allergen Cry j 1 which stimulate an immune  
 CC response in dogs. The invention also encompasses DNA encoding the Cry j 1  
 CC T-cell epitopes, a vector comprising a Cry j 1 T-cell epitope-encoding  
 CC DNA, and DNA vaccine containing a vector of the invention for the  
 CC treatment of allergic dermatitis. The Cry j 1 T-cell epitopes and their  
 CC encoding nucleic acids are useful in the preparation of vaccines,  
 CC especially DNA vaccines, for the treatment of allergic dermatitis in  
 CC dogs. The DNA vaccine of the invention can be prepared inexpensively with  
 CC easier manufacturing and purification processes. The present sequence  
 CC represents a specifically claimed Cry j 1 T-cell epitope of the  
 CC invention.

XX Sequence 20 AA;

Query Match 88.9%; Score 32; DB 7; Length 20;  
 Best Local Similarity 100.0%; Pred. No. 2.5e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 OY 2 GGPCV 6  
 |||||  
 Db 14 GGPCV 18

RESULT 26

AA45579 standard; protein; 30 AA.

AC AA45579;

DT 25-MAR-2003 (revised)

DT 13-JUL-1994 (first entry)

DE Cry j 1 pollen allergen peptide CJI-41.

KW Japanese cedar; detection; allergy; treatment; diagnosis; T cell epitope;  
 KM sensitivity.

OS Cryptomeria japonica.

PN WO9401560-A1.

PD 20-JAN-1994.

PF 15-JAN-1993; 93MO-US000139.

PR 10-JUL-1992; 92MO-US005661.

PR 01-SEP-1992; 92US-00938990.

XX (IMMU-) IMMUNOLOGIC PHARM CORP.

XX Griffith J, Pollock J, Bond JF, Garman RD, Kuo M;

XX WPI; 1994-035066/04.

PT Antigens derived from Japanese cedar pollen allergen Cry j 1 - contain at  
 PT least two T cell epitope(s), used to treat or diagnose allergy.

PS Claim 76; Fig 18; 137pp; English.

CC The sequence is that of an isolated peptide of the Japanese cedar pollen  
 CC allergen Cry j 1. The peptide, CJI-41, can be used for the treatment and  
 CC diagnosis of allergies associated with Japanese cedar pollen. It has  
 CC enhanced therapeutic properties but reduced side effects compared to  
 CC naturally occurring allergens. (Updated on 25-MAR-2003 to correct PN  
 CC field.)

XX Sequence 30 AA;

Query Match 88.9%; Score 32; DB 2; Length 30;

Best Local Similarity 100.0%; Pred. No. 3.5e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GGPCV 6  
 |||||  
 Db 24 GGPCV 28

RESULT 27

AA44683 standard; peptide; 30 AA.

AC AA44683;

DT 01-MAY-1998 (first entry)

XX T-cell epitope peptide #2 of sugi pollen antigen.

XX T-cell epitope, sugi pollen antigen; sugi pollinosis.

OS Synthetic.  
OS Cryptomeria japonica.  
XX  
XX JP10007700-A.  
XX  
XX 13-JAN-1998.  
XX  
XX 24-JUN-1996; 96GP-00163287.  
XX  
XX 24-JUN-1996; 96GP-00163287.  
XX  
XX (MEIJ ) MEIJ SEIKA KAISHA LTD.  
XX (DAIL ) DAICEL CHEM IND LTD.  
XX  
XX WPI; 1998-133630/13.  
XX  
XX  
XX T cell epitope peptide of sugi pollen antigen - useful in the treatment  
XX of sugi pollinosis.  
XX  
XX Claim 1; Page 4; 14pp; Japanese.  
XX  
XX T-cell epitope peptides AAM4682-88 and their derivatives react with sugi  
XX pollinosis patient peripheral blood T lymphocytes. A composition prepared  
XX by combining at least 2 of the above peptides and/or their derivatives is  
XX used for the prevention and treatment of sugi pollinosis  
XX  
XX Sequence 30 AA;  
SQ  
Query Match 88.9%; Score 32; DB 2; Length 30;  
Best Local Similarity 100.0%; Pred. NO. 3.5e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
CY 2 GGPCV 6  
Db 4 GGPCV 8  
RESULT 28  
ADL1905  
ID ADL1905 standard; protein; 33 AA.  
XX  
XX ADL1905;  
AC  
XX 06-MAY-2004 (first entry)  
DT  
XX Ravenoxin-II protein.  
DE  
XX Bacillus thuringiensis; spider toxin gene; biopesticide.  
XX  
XX Bacillus thuringiensis.  
OS  
XX CN1366822-A.  
XX  
XX 04-SEP-2002.  
PD  
XX 18-JUL-2001; 2001CN-00114592.  
PF  
XX 18-JUL-2001; 2001CN-00114592.  
PR  
XX (UYHU-) UNIV HUNAN NORMAL.  
XX  
XX Xia L, Liang S, Ding X;  
PI  
XX WPI; 2003-483110/46.  
DR  
XX N-Psdb; ADL1906.  
DR  
XX  
XX New strain of Bacillus thuringiensis, containing a spider toxin gene and  
XX a promoter sequence, is used as a biopesticide.  
XX  
XX Disclosure; SEQ ID NO 9; 28pp; Chinese.  
XX  
XX The present invention relates to a Bacillus thuringiensis strain  
XX comprising a spider toxin gene and a strong promoter sequence. The

CC B.thuringiensis is used as a biopesticide as it can produce the  
CC B.thuringiensis toxin and a spider toxin. The present sequence represents  
CC Ravenoxin-II protein.  
XX  
XX Sequence 33 AA;  
SQ  
Query Match 88.9%; Score 32; DB 7; Length 33;  
Best Local Similarity 100.0%; Pred. NO. 3.8e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
CY 2 GGPCV 6  
Db 6 GGPCV 10  
RESULT 29  
AAM91570  
ID AAM91570 standard; protein; 43 AA.  
XX  
XX AAM91570;  
AC  
XX  
XX 07-NOV-2001 (first entry)  
DT  
XX  
XX Human immune/haematopoietic antigen SEQ ID NO:19163.  
XX  
XX Human immune/haematopoietic antigen  
XX  
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
XX cytostatic; gene therapy; vaccine; metastasis.  
XX  
XX Homo sapiens.  
OS  
XX WO200157182-A2.  
XX  
XX 09-AUG-2001.  
PD  
XX  
XX 17-JAN-2001; 2001WO-US001354.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
XX  
XX 04-FEB-2000; 2000US-0180628P.  
XX  
XX 24-FEB-2000; 2000US-0184664P.  
XX  
XX 02-MAR-2000; 2000US-0186350P.  
XX  
XX 16-MAR-2000; 2000US-0189874P.  
XX  
XX 17-MAR-2000; 2000US-0190076P.  
XX  
XX 18-APR-2000; 2000US-0198123P.  
XX  
XX 19-MAY-2000; 2000US-0205151P.  
XX  
XX 07-JUN-2000; 2000US-0209467P.  
XX  
XX 28-JUN-2000; 2000US-0214866P.  
XX  
XX 30-JUN-2000; 2000US-0215135P.  
XX  
XX 07-JUL-2000; 2000US-0216647P.  
XX  
XX 07-JUL-2000; 2000US-0216880P.  
XX  
XX 11-JUL-2000; 2000US-0217487P.  
XX  
XX 11-JUL-2000; 2000US-0217496P.  
XX  
XX 14-JUL-2000; 2000US-0218290P.  
XX  
XX 26-JUL-2000; 2000US-0220963P.  
XX  
XX 26-JUL-2000; 2000US-0220964P.  
XX  
XX 14-AUG-2000; 2000US-0224518P.  
XX  
XX 14-AUG-2000; 2000US-0224519P.  
XX  
XX 14-AUG-2000; 2000US-0225213P.  
XX  
XX 14-AUG-2000; 2000US-0225214P.  
XX  
XX 14-AUG-2000; 2000US-0225266P.  
XX  
XX 14-AUG-2000; 2000US-0225267P.  
XX  
XX 14-AUG-2000; 2000US-0225268P.  
XX  
XX 14-AUG-2000; 2000US-0225270P.  
XX  
XX 14-AUG-2000; 2000US-0225447P.  
XX  
XX 14-AUG-2000; 2000US-0225757P.  
XX  
XX 14-AUG-2000; 2000US-0225758P.  
XX  
XX 14-AUG-2000; 2000US-0225759P.  
XX  
XX 18-AUG-2000; 2000US-0226279P.  
XX  
XX 22-AUG-2000; 2000US-0226681P.  
XX  
XX 22-AUG-2000; 2000US-0226868P.  
XX  
XX 22-AUG-2000; 2000US-0227182P.  
XX  
XX 23-AUG-2000; 2000US-0227189P.  
XX  
XX 30-AUG-2000; 2000US-0228924P.  
XX  
XX 01-SEP-2000; 2000US-0229287P.  
XX

PR 01-SEP-2000; 2000US-0229343P.  
 PR 01-SEP-2000; 2000US-0229344P.  
 PR 01-SEP-2000; 2000US-0229345P.  
 PR 05-SEP-2000; 2000US-0229503P.  
 PR 05-SEP-2000; 2000US-0229513P.  
 PR 06-SEP-2000; 2000US-0230437P.  
 PR 06-SEP-2000; 2000US-0230438P.  
 PR 08-SEP-2000; 2000US-0231242P.  
 PR 08-SEP-2000; 2000US-0231243P.  
 PR 08-SEP-2000; 2000US-0231244P.  
 PR 08-SEP-2000; 2000US-0231413P.  
 PR 08-SEP-2000; 2000US-0231414P.  
 PR 08-SEP-2000; 2000US-0232080P.  
 PR 12-SEP-2000; 2000US-0231968P.  
 PR 12-SEP-2000; 2000US-0231969P.  
 PR 14-SEP-2000; 2000US-0233397P.  
 PR 14-SEP-2000; 2000US-0233398P.  
 PR 14-SEP-2000; 2000US-0233399P.  
 PR 14-SEP-2000; 2000US-0234001P.  
 PR 14-SEP-2000; 2000US-0234011P.  
 PR 14-SEP-2000; 2000US-0233063P.  
 PR 14-SEP-2000; 2000US-0233064P.  
 PR 14-SEP-2000; 2000US-0233065P.  
 PR 21-SEP-2000; 2000US-0234223P.  
 PR 21-SEP-2000; 2000US-0234274P.  
 PR 25-SEP-2000; 2000US-0234997P.  
 PR 25-SEP-2000; 2000US-0234998P.  
 PR 26-SEP-2000; 2000US-0235484P.  
 PR 27-SEP-2000; 2000US-0235834P.  
 PR 27-SEP-2000; 2000US-0235835P.  
 PR 29-SEP-2000; 2000US-0236327P.  
 PR 29-SEP-2000; 2000US-0236367P.  
 PR 29-SEP-2000; 2000US-0236368P.  
 PR 29-SEP-2000; 2000US-0236369P.  
 PR 29-SEP-2000; 2000US-0236370P.  
 PR 02-OCT-2000; 2000US-0236802P.  
 PR 02-OCT-2000; 2000US-0237037P.  
 PR 02-OCT-2000; 2000US-0237038P.  
 PR 02-OCT-2000; 2000US-0237039P.  
 PR 02-OCT-2000; 2000US-0237040P.  
 PR 13-OCT-2000; 2000US-0239335P.  
 PR 13-OCT-2000; 2000US-0239337P.  
 PR 20-OCT-2000; 2000US-0240960P.  
 PR 20-OCT-2000; 2000US-0241221P.  
 PR 20-OCT-2000; 2000US-0241785P.  
 PR 20-OCT-2000; 2000US-0241786P.  
 PR 20-OCT-2000; 2000US-0241808P.  
 PR 20-OCT-2000; 2000US-0241809P.  
 PR 20-OCT-2000; 2000US-0241825P.  
 PR 01-NOV-2000; 2000US-0244617P.  
 PR 08-NOV-2000; 2000US-0246474P.  
 PR 08-NOV-2000; 2000US-0246475P.  
 PR 08-NOV-2000; 2000US-0246476P.  
 PR 08-NOV-2000; 2000US-0246477P.  
 PR 08-NOV-2000; 2000US-0246478P.  
 PR 08-NOV-2000; 2000US-0246523P.  
 PR 08-NOV-2000; 2000US-0246524P.  
 PR 08-NOV-2000; 2000US-0246525P.  
 PR 08-NOV-2000; 2000US-0246526P.  
 PR 08-NOV-2000; 2000US-0246527P.  
 PR 08-NOV-2000; 2000US-0246528P.  
 PR 08-NOV-2000; 2000US-0246532P.  
 PR 08-NOV-2000; 2000US-0246610P.  
 PR 08-NOV-2000; 2000US-0246611P.  
 PR 08-NOV-2000; 2000US-0246613P.  
 PR 17-NOV-2000; 2000US-0249207P.  
 PR 17-NOV-2000; 2000US-0249208P.  
 PR 17-NOV-2000; 2000US-0249209P.  
 PR 17-NOV-2000; 2000US-0249210P.  
 PR 17-NOV-2000; 2000US-0249211P.  
 PR 17-NOV-2000; 2000US-0249212P.

PR 17-NOV-2000; 2000US-0249213P.  
 PR 17-NOV-2000; 2000US-0249214P.  
 PR 17-NOV-2000; 2000US-0249215P.  
 PR 17-NOV-2000; 2000US-0249216P.  
 PR 17-NOV-2000; 2000US-0249217P.  
 PR 17-NOV-2000; 2000US-0249218P.  
 PR 17-NOV-2000; 2000US-0249244P.  
 PR 17-NOV-2000; 2000US-0249245P.  
 PR 17-NOV-2000; 2000US-0249264P.  
 PR 17-NOV-2000; 2000US-0249265P.  
 PR 17-NOV-2000; 2000US-0249297P.  
 PR 17-NOV-2000; 2000US-0249299P.  
 PR 17-NOV-2000; 2000US-0249300P.  
 PR 01-DEC-2000; 2000US-0250160P.  
 PR 01-DEC-2000; 2000US-0250391P.  
 PR 05-DEC-2000; 2000US-0251030P.  
 PR 05-DEC-2000; 2000US-0251988P.  
 PR 05-DEC-2000; 2000US-0256719P.  
 PR 06-DEC-2000; 2000US-0251479P.  
 PR 08-DEC-2000; 2000US-0251856P.  
 PR 08-DEC-2000; 2000US-0251858P.  
 PR 08-DEC-2000; 2000US-0251989P.  
 PR 08-DEC-2000; 2000US-0251990P.  
 PR 11-DEC-2000; 2000US-0254097P.  
 PR 05-JAN-2001; 2001US-0259678P.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 PA  
 XX  
 XX Rosen CA, Barash SC, Ruben SM;  
 PI  
 XX MPI; 2001-483426/52.  
 XX N-PSDB; AAK64351.  
 DR  
 XX  
 PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
 useful for preventing, diagnosing and/or treating cancers and metastasis.  
 XX  
 PS Claim 11; SEQ ID NO 19163; 3071pp + Sequence listing; English.  
 XX  
 CC AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)  
 CC amino acid sequences given in AAK62170 to AAK91921. (I) have cytostatic  
 CC activity, and can be used in gene therapy and vaccine production. (I)  
 CC proteins and polynucleotides may be used in the prevention, diagnosis and  
 CC treatment of diseases associated with inappropriate (I) expression. For  
 CC example, they may be used to treat disorders associated with decreased  
 CC expression by rectifying mutations or deletions in a patient's genome  
 CC that affect the activity of (I) by expressing inactive proteins or to  
 CC supplement the patient's own production of (I). Additionally, (I)  
 CC polynucleotides may be used to produce the secreted (I), by inserting the  
 CC nucleic acids into a host cell and culturing the cell to express the  
 CC protein. (I) proteins and polynucleotides may be used to prevent,  
 CC diagnose and treat immune/haematopoietic-related diseases, especially  
 CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703  
 CC to AAK67694 represent human immune/hematopoietic antigen genomic  
 CC sequences from the present invention. AAK54942 to AAK54950 and AAK62169  
 CC represent sequences used in the exemplification of the present invention  
 XX  
 SQ Sequence 43 AA;  
 Query Match 88.9%; Score 32; DB 4; Length 43;  
 Best Local Similarity 100.0%; Pred. No. 4.8e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 GGCPCV 6  
 |||||  
 Db 2 GGCPCV 6  
 RESULT 30  
 ABO57709  
 ID ABO57709 standard; protein; 48 AA.  
 XX  
 AC ABO57709;

XX 29-JUL-2004 (first entry)  
XX Human genome derived single exon protein #3943.  
XX Human; gene expression; single exon probe; microarray;  
XX alternative splicing event; genomic alteration.  
XX Homo sapiens.  
XX US2003194704-A1.  
XX 16-OCT-2003.  
XX 03-APR-2002; 2002US-00029386.  
XX 03-APR-2002; 2002US-00029386.  
XX 03-APR-2002; 2002US-00029386.  
XX (PENN/) PENN S G.  
XX (RANK/) RANK D R.  
XX (HANZ/) HANZEL D K.  
XX Penn SG, Rank DR, Hanzel DK,  
XX WPI, 2004-119264/12.  
XX New human genome-derived single exon nucleic acid probes useful for human  
XX gene expression analysis, for identifying or characterizing alternative  
XX splicing events, for assessing genomic alterations or as tools for  
XX surveying tissues.  
XX Claim 45; SEQ ID NO 31343; 80pp; English.  
XX The invention relates to a nucleic acid probe for measuring human gene  
XX expression, comprising any of the 27,400 fully defined nucleotide  
XX sequences in the specification, or their complements or fragments, and  
XX encoding at least 8 amino acids of any of the 6888 amino acid sequences  
XX fully defined in the specification. The probe is a single exon probe that  
XX hybridizes under high stringency conditions to a nucleic acid molecule  
XX expressed in human cells or tissues. Also included are a spatially-  
XX addressable set of single exon nucleic acid probes for measuring human  
XX gene expression (comprising a plurality of single exon nucleic acid  
XX probes cited above, where each of the plurality of probes is separately  
XX and addressably isolatable or amplifiable from the plurality), a single  
XX exon microarray for measuring human gene expression, a method of  
XX measuring human gene expression, a vector comprising the single exon  
XX probe cited above, an ORF-encoded peptide comprising at least 8  
XX contiguous amino acids of any of the above-mentioned amino acid  
XX sequences (optionally with conservative amino acid substitutions), an  
XX isolated antibody that binds specifically to a peptide cited above, a  
XX method of selling and/or licensing single exon probes or microarrays to  
XX a customer desiring to measure gene expression, a method of providing  
XX human gene expression data by subcription, and a computer-readable  
XX storage medium which contains a database having a plurality of records  
XX (each record including data on the expression of a single exon probe  
XX cited above. The probe, methods and apparatus are useful in gene  
XX expression analysis. The probes may be used as tools for surveying  
XX tissues to detect the presence of expressed messages that contain their  
XX specific exon, or in constructing genome-derived single exon microarrays.  
XX In addition, the probes are used in identifying and characterizing  
XX alternative splicing events, in detecting and characterizing gross  
XX alterations in the genomic locus that includes their exon, in assessing  
XX smaller genomic alterations, in priming the synthesis of nucleic acids,  
XX or in expressing the ORF-encoded peptide. The present sequence is a human  
XX single exon probe protein of the invention. Note: The sequence data for  
XX this patent did not form part of the printed specification, but was  
XX obtained in electronic format directly from USPTO at  
XX seqdata.uspto.gov/sequence.html?DocID=20030194704  
XX  
XX Sequence 48 AA:  
SQ

Query Match 88.9%; Score 32; DB 8; Length 48;  
Best Local Similarity 100.0%; Pred. No. 5.2e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2 GPCV 6  
DB 39 GPCV 43  
RESULT 31  
AAM82764  
ID AAM82764 standard; protein, 53 AA.  
XX  
XX AAM82764;  
XX  
XX 07-NOV-2001 (first entry)  
XX Human immune/haematopoietic antigen SEQ ID NO:10357.  
XX  
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
XX cytostatic; gene therapy; vaccine; metastasis.  
XX Homo sapiens.  
XX WO200157182-A2.  
XX  
XX 09-AUG-2001.  
XX  
XX 17-JAN-2001; 2001WO-US001354.  
XX  
XX 31-JAN-2000; 2000US-0179065P.  
XX 04-FEB-2000; 2000US-0180628P.  
XX 24-FEB-2000; 2000US-0184664P.  
XX 02-MAR-2000; 2000US-0186350P.  
XX 16-MAR-2000; 2000US-0189874P.  
XX 17-MAR-2000; 2000US-0190076P.  
XX 18-APR-2000; 2000US-0198123P.  
XX 19-MAY-2000; 2000US-0205151P.  
XX 07-JUN-2000; 2000US-0209467P.  
XX 28-JUN-2000; 2000US-0214886P.  
XX 30-JUN-2000; 2000US-0215135P.  
XX 07-JUL-2000; 2000US-0216647P.  
XX 07-JUL-2000; 2000US-0216880P.  
XX 11-JUL-2000; 2000US-0217487P.  
XX 11-JUL-2000; 2000US-0217496P.  
XX 14-JUL-2000; 2000US-0218290P.  
XX 26-JUL-2000; 2000US-0220963P.  
XX 26-JUL-2000; 2000US-0220964P.  
XX 14-AUG-2000; 2000US-0224518P.  
XX 14-AUG-2000; 2000US-0224519P.  
XX 14-AUG-2000; 2000US-0225213P.  
XX 14-AUG-2000; 2000US-0225214P.  
XX 14-AUG-2000; 2000US-0225265P.  
XX 14-AUG-2000; 2000US-0225267P.  
XX 14-AUG-2000; 2000US-0225268P.  
XX 14-AUG-2000; 2000US-0225270P.  
XX 14-AUG-2000; 2000US-0225447P.  
XX 14-AUG-2000; 2000US-0225757P.  
XX 14-AUG-2000; 2000US-0225758P.  
XX 14-AUG-2000; 2000US-0225759P.  
XX 14-AUG-2000; 2000US-0226279P.  
XX 22-AUG-2000; 2000US-0226811P.  
XX 22-AUG-2000; 2000US-0226868P.  
XX 22-AUG-2000; 2000US-0227182P.  
XX 23-AUG-2000; 2000US-0227009P.  
XX 30-AUG-2000; 2000US-0228924P.  
XX 01-SEP-2000; 2000US-0229287P.  
XX 01-SEP-2000; 2000US-0229343P.  
XX 01-SEP-2000; 2000US-0229344P.  
XX 01-SEP-2000; 2000US-0229345P.  
XX 09-SEP-2000; 2000US-0229309P.  
XX 05-SEP-2000; 2000US-0229513P.  
XX 06-SEP-2000; 2000US-0230437P.  
XX 06-SEP-2000; 2000US-0230438P.  
XX 08-SEP-2000; 2000US-0231242P.

PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 25-SEP-2000; 2000US-0235484P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 27-SEP-2000; 2000US-0235835P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239933P.  
PR 13-OCT-2000; 2000US-0239937P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241783P.  
PR 20-OCT-2000; 2000US-0241786P.  
PR 20-OCT-2000; 2000US-0241787P.  
PR 20-OCT-2000; 2000US-0241808P.  
PR 20-OCT-2000; 2000US-0241809P.  
PR 20-OCT-2000; 2000US-0241826P.  
PR 01-NOV-2000; 2000US-0244617P.  
PR 08-NOV-2000; 2000US-0246473P.  
PR 08-NOV-2000; 2000US-0246475P.  
PR 08-NOV-2000; 2000US-0246476P.  
PR 08-NOV-2000; 2000US-0246477P.  
PR 08-NOV-2000; 2000US-0246478P.  
PR 08-NOV-2000; 2000US-0246523P.  
PR 08-NOV-2000; 2000US-0246524P.  
PR 08-NOV-2000; 2000US-0246525P.  
PR 08-NOV-2000; 2000US-0246526P.  
PR 08-NOV-2000; 2000US-0246527P.  
PR 08-NOV-2000; 2000US-0246528P.  
PR 08-NOV-2000; 2000US-0246532P.  
PR 08-NOV-2000; 2000US-0246603P.  
PR 08-NOV-2000; 2000US-0246610P.  
PR 08-NOV-2000; 2000US-0246611P.  
PR 08-NOV-2000; 2000US-0246613P.  
PR 17-NOV-2000; 2000US-0249207P.  
PR 17-NOV-2000; 2000US-0249208P.  
PR 17-NOV-2000; 2000US-0249209P.  
PR 17-NOV-2000; 2000US-0249210P.  
PR 17-NOV-2000; 2000US-0249211P.  
PR 17-NOV-2000; 2000US-0249212P.  
PR 17-NOV-2000; 2000US-0249213P.  
PR 17-NOV-2000; 2000US-0249214P.  
PR 17-NOV-2000; 2000US-0249215P.  
PR 17-NOV-2000; 2000US-0249216P.  
PR 17-NOV-2000; 2000US-0249217P.  
PR 17-NOV-2000; 2000US-0249218P.  
PR 17-NOV-2000; 2000US-0249244P.  
PR 17-NOV-2000; 2000US-0249245P.

PR 17-NOV-2000; 2000US-0249264P.  
PR 17-NOV-2000; 2000US-0249265P.  
PR 17-NOV-2000; 2000US-0249297P.  
PR 17-NOV-2000; 2000US-0249299P.  
PR 17-NOV-2000; 2000US-0249300P.  
PR 01-DEC-2000; 2000US-0250160P.  
PR 01-DEC-2000; 2000US-0250391P.  
PR 05-DEC-2000; 2000US-0251030P.  
PR 05-DEC-2000; 2000US-0251988P.  
PR 05-DEC-2000; 2000US-0256719P.  
PR 06-DEC-2000; 2000US-0251479P.  
PR 08-DEC-2000; 2000US-0251856P.  
PR 08-DEC-2000; 2000US-0251868P.  
PR 08-DEC-2000; 2000US-0251869P.  
PR 08-DEC-2000; 2000US-0251989P.  
PR 08-DEC-2000; 2000US-0251990P.  
PR 11-DEC-2000; 2000US-0254097P.  
PR 05-JAN-2001; 2001US-0259678P.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
PA  
XX  
PI Rosen CA, Barash SC, Ruben SM;  
XX  
XX WPI; 2001-483426/52.  
DR N-PSDB; AAK55545.  
XX  
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
PT useful for preventing, diagnosing and/or treating cancers and metastasis.  
XX  
XX Claim 11; SEQ ID NO 10357; 3071pp + Sequence Listing; English.  
PS  
XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)  
XX amino acid sequences given in AAM82170 to AAM91921. (I) have cytosolic  
XX activity, and can be used in gene therapy and vaccine production. (I)  
XX proteins and polynucleotides may be used in the prevention, diagnosis and  
XX treatment of diseases associated with inappropriate (I) expression. For  
XX example, they may be used to treat disorders associated with decreased  
XX expression by rectifying mutations or deletions in a patient's genome  
XX that affect the activity of (I) by expressing inactive proteins or to  
XX supplement the patient's own production of (I). Additionally, (I)  
XX polynucleotides may be used to produce the secreted (I), by inserting the  
XX nucleic acids into a host cell and culturing the cell to express the  
XX protein. (I) proteins and polynucleotides may be used to prevent,  
XX diagnose and treat immune/hematopoietic-related diseases, especially  
XX cancers and cancer metastases of hematopoietic-derived cells. AAK64703  
XX to AAK87694 represent human immune/hematopoietic antigen genomic  
XX sequences from the present invention. AAK54942 to AAK54950 and AAM82169  
XX represent sequences used in the exemplification of the present invention  
XX  
SQ Sequence 53 AA;  
Query Match 88.9%; Score 32; DB 4; Length 53;  
Best Local Similarity 100.0%; Pred. No. 5.7e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
CY 1 AGSPC 5  
DB 39 AGSPC 43  
RESULT 32  
AAW05836  
ID AAW05836 standard; protein; 54 AA.  
AC AAW05836;  
XX  
XX 25-MAR-2003 (revised)  
DT 28-JAN-1997 (first entry)  
XX  
XX Mouse M-Serrate-1 (DSL domain upstream region).  
DB M-Serrate-1; Notch; cell differentiation; cell fate; tissue repair;  
XX M-Serrate-1; Notch; cell differentiation; cell fate; tissue repair;  
XX central nervous system; cancer; therapy; diagnosis.  
KW

```

XX OS Mus sp.
XX PN WO9627610-A1.
XX PD 12-SEP-1996.
XX PF 07-MAR-1996; 96WO-US003172.
XX PR 07-MAR-1995; 95US-00400159.
XX PA (UYTA ) UNIV YALE.
XX PA (IMCR ) IMPERIAL CANCER RES TECHNOLOGY.
XX PI Ishorowicz D, Henrique DMP, Lewis JH, Myat AM;
XX PI Artavanis-Tsakonas S, Mann RS, Gray GE;
XX DR WPI; 1996-425379/42.
XX DR N-PSDB; AA140095.
XX PT Vertebrate Serrate protein and related DNA - used to treat or prevent
XX PT malignancies characterized by increased Notch activity.
XX PS Disclosure; Page 65; 161pp; English.
XX CC A polypeptide (AAW05836) corresponds to an N-terminal sequence of mouse
XX CC Serrate homologue. M-Serrate-1, located upstream of the DSL domain. It is
XX CC the product of the 5' end (AA140095) of a cDNA clone isolated from mouse
XX CC 10.5-day embryo. Another polypeptide (AAW05837) located within the
XX CC epidermal growth factor-like repeats of Serrate was also identified. M-
XX CC Serrate-1 is expressed in the central nervous system, peripheral nervous
XX CC system, limb, kidney, lens and vascular system. It is a ligand for Notch
XX CC and is thought to play an important role in determining cell fate, esp.
XX CC in the central nervous system. (Updated on 25-MAR-2003 to correct PI
XX CC field.)
XX SQ Sequence 54 AA;

Query Match      88.9%; Score 32; DB 2; Length 54;
Best Local Similarity 100.0%; Pred. No. 5.8e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGPG 5
   |||||
Db 5 AGGPG 9

RESULT 33
AAVS9595
ID AAVS9595 standard; protein; 54 AA.
XX AC AAVS9595;
XX DT 05-APR-2000 (first entry)
XX DE Mouse Serrate protein sequence.
XX KW Serrate; cancerous condition; melanoma; lung cancer; breast cancer;
XX KW nervous system disorder; infection; nutritional disease; therapy;
XX KW cell proliferation promoter; tissue regeneration; mouse.
XX OS Mus sp.
XX PN US6004924-A.
XX PD 21-DEC-1999.
XX PF 06-MAR-1996; 96US-00611729.
XX PR 11-DEC-1991; 91US-00808458.
XX PR 14-SEP-1993; 93US-00121979.
XX PR 07-JUN-1994; 94US-00255102.
XX PR 07-MAR-1995; 95US-00400159.

```

```

XX PA (IMCR ) IMPERIAL CANCER RES TECHNOLOGY.
XX PA (UYTA ) UNIV YALE.
XX PI Ish-Horowicz D, Henrique DMP, Myat AM, Fleming RJ;
XX PI Artavanis-Tsakonas S, Gray GE, Mann RS, Lewis JH;
XX DR WPI; 2000-105089/09.
XX DR N-PSDB; AA249094.
XX PT Purified Serrate proteins useful for treating neoplasias, nervous
XX PT disorders and for promoting cell proliferation and tissue regeneration
XX PT and repair.
XX PS Claim 1, Col 135-136; 114pp; English.
XX CC This sequence represents a mouse serrate protein. The invention relates
XX CC to purified vertebrate (mouse, chick, and human) serrate proteins. The
XX CC Serrate proteins may be administered to treat a cancerous condition (e.g.
XX CC melanoma, lung cancer and breast cancer) by preventing progression from a
XX CC preneoplastic or nonmalignant state into a neoplastic or malignant state.
XX CC It may also be used to treat nervous system disorders (such as lesions
XX CC caused by infections, nutritional disease and toxic substances) and to
XX CC promote cell proliferation and tissue regeneration and repair. The
XX CC protein itself is administered to supplement a patient's own production
XX CC of Serrate proteins (if levels of expression are low) or to compensate
XX CC for expression of inactive proteins due to genetic mutations. The protein
XX CC may also be used in the production of antibodies against Serrate proteins
XX CC which may be used to either down regulate Serrate activity or to detect
XX CC Serrate proteins in samples (for example via enzyme-linked immunosorbent
XX CC assay (ELISA)). The proteins may also be used to study Serrate expression
XX CC and its role in metabolism and to assay for agents which modulate its
XX CC expression and activity
XX SQ Sequence 54 AA;

Query Match      88.9%; Score 32; DB 3; Length 54;
Best Local Similarity 100.0%; Pred. No. 5.8e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGPG 5
   |||||
Db 5 AGGPG 9

RESULT 34
ADL71374
ID ADL71374 standard; peptide; 54 AA.
XX AC ADL71374;
XX DT 20-MAY-2004 (first entry)
XX DE Mouse serrate-1 partial protein #1.
XX KW Serrate protein; notch protein; breast cancer; fibrosarcoma; myxosarcoma;
XX KW nervous system disorder; motor neuron disorder; cirrhosis of liver; keloid; psoriasis;
XX KW benign dysproliferative disorder; carcinoma of liver; hepatocellular;
XX KW Chediak-Higashi syndrome; cytostatic; hepatocellular; vulnary;
XX KW antiproliferative; immunomodulator; serrate-1; mouse.
XX OS Mus sp.
XX PN US6703489-B1.
XX PD 09-MAR-2004.
XX PF 19-NOV-1998; 98US-00195524.
XX PR 07-MAR-1995; 95US-00400159.
XX PR 06-MAR-1996; 96US-00611729.
XX PA (UYTA ) UNIV YALE.

```

PA (IMCR ) IMPERIAL CANCER RES TECHNOLOGY LTD.  
 XX ICh-Horowicz D, Henrique DMP, Lewis JH, Myat AM, Fleming RJ;  
 PI Attavanis-Tsakonas S, Mann RS, Gray GB;  
 XX WPI; 2004-224050/21.  
 DR N-PSDB; ADL71373.  
 XX  
 PT New antibody binding to first Serrate protein or human Serrate protein,  
 PT useful for treating nervous system disorders, malignancies, cirrhosis of  
 PT liver.  
 PS Disclosure; SEQ ID NO 16; 117pp; English.  
 XX  
 CC The present invention relates to novel serrate proteins which bind a  
 CC notch protein and polynucleotides encoding such proteins. Sequences of  
 CC the invention are useful for treating malignancies and related disorders  
 CC such as breast cancer, fibrosarcoma, myosarcoma, etc., nervous system  
 CC disorders, motor neuron disorders, benign dysplastic disorders,  
 CC cirrhosis of liver, keloid, psoriasis, Chediak-Higashi syndrome etc. They  
 CC are also useful for diagnosing diseases and disorders related to the  
 CC expression of Serrate protein. The present sequence is mouse serrate-1  
 CC partial protein used in the exemplification of the invention.  
 CC  
 SQ Sequence 54 AA;  
 Query Match 88.9%; Score 32; DB 8; Length 54;  
 Best Local Similarity 100.0%; Pred. No. 5.8e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 1 AGGPC 5  
 Db 5 AGGPC 9  
 RESULT 35  
 ADK34982  
 ID ADK34982 standard; protein; 71 AA.  
 AC ADK34982;  
 XX  
 DT 06-MAY-2004 (first entry)  
 XX  
 DE Novel human polypeptide SegID7064.  
 XX  
 KW antiarthritic; antiparkinsonian; neuroprotective; nootropic;  
 KW immunosuppressive; cytosstatic; antiposoriatic; antiinflammatory;  
 KW antibacterial; antiviral; antifungal; antiparasitic; gene therapy;  
 KW arthritis; Parkinson's; Alzheimer's; autoimmune disease; cancer;  
 KW psoriasis; inflammatory bowel disease; infection; bacteria; virus;  
 KW fungus; parasite; human.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key location/Qualifiers  
 FT Misc-difference 1..71  
 FT /label= OTHER  
 FT /note= "OTHER= All Xaa's in this sequence are unknown  
 FT amino acids or the site of a stop codon within the DNA  
 FT sequence"  
 XX  
 EN WO200216439-A2.  
 XX  
 PD 28-FEB-2002.  
 XX  
 PF 05-MAR-2001; 2001WO-US004941.  
 XX  
 PR 07-MAR-2000; 2000US-00519705.  
 PR 19-MAY-2000; 2000US-00574454.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Tang YT, Liu C, Dimanac RT;

XX WPI; 2002-280916/32.  
 DR  
 XX Isolated polynucleotide encoding bone marrow derived polypeptides useful  
 PT for treating, e.g., Parkinson's, Alzheimer's, cancer, arthritis, Crohn's  
 PT disease, and inflammatory bowel disease.  
 XX  
 PS Claim 20; SEQ ID NO 7064; 504pp; English.  
 XX  
 CC This invention relates to a novel isolated polynucleotide comprising a  
 CC nucleotide sequence selected from one of 1680 sequences, a mature protein  
 CC coding portion of them, an active domain of them and their complementary  
 CC sequences. The invention may be useful for the production of compounds  
 CC with an antiarthritic, antiparkinsonian, neuroprotective, nootropic,  
 CC immunosuppressive, cytosstatic, antiposoriatic, antiinflammatory,  
 CC antibacterial, antiviral, antifungal or antiparasitic activity. In  
 CC addition, the disclosed sequences may be useful for gene therapy. The  
 CC polypeptides or their antibodies are useful for treating many diseases  
 CC such as arthritis, Parkinson's, Alzheimer's, autoimmune diseases, cancer,  
 CC psoriasis, inflammatory bowel disease and infections caused by bacteria,  
 CC viruses, fungi or parasites. The present sequence is that of a human  
 CC polypeptide of the invention.  
 CC  
 SQ Sequence 71 AA;  
 Query Match 88.9%; Score 32; DB 5; Length 71;  
 Best Local Similarity 100.0%; Pred. No. 7.3e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 2 GGPCV 6  
 Db 28 GGPCV 32  
 RESULT 36  
 AAU65187  
 ID AAU65187 standard; protein; 74 AA.  
 AC AAU65187;  
 XX  
 DT 27-FEB-2002 (first entry)  
 XX  
 DE Propionibacterium acnes immunogenic protein #26083.  
 XX  
 KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
 KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
 KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
 KW dermatological; osteopethic; neuroprotectant.  
 XX  
 OS Propionibacterium acnes.  
 XX  
 PN WO200181581-A2.  
 XX  
 PD 01-NOV-2001.  
 XX  
 PF 20-APR-2001; 2001WO-US012865.  
 XX  
 PR 21-APR-2000; 2000US-0199047P.  
 PR 02-JUN-2000; 2000US-020841P.  
 PR 07-JUL-2000; 2000US-0216747P.  
 XX  
 PA (CORI-) CORIXA CORP.  
 XX  
 PI Skelky YAM, Persing DH, Mitcham JL, Wang SS, Bhacia A;  
 PI L'maisonneuve J, Zhang Y, Jen S, Carter D;  
 XX  
 DR WPI; 2001-616774/71.  
 DR N-PSDB; AAS59661.  
 XX  
 PT Propionibacterium acnes polypeptides and nucleic acids useful for  
 PT vaccinating against and diagnosing infections, especially useful for  
 PT treating acne vulgaris.

PS Example 1, SEQ ID NO 26382, 1063pp; English.  
XX  
CC Sequences AAU39105-AAU68017 represent *Propionibacterium acnes* immunogenic  
CC polypeptides. The proteins and their associated DNA sequences are used in  
CC the treatment, prevention and diagnosis of medical conditions caused by  
CC *P. acnes*. The disorders include SAPHO syndrome (synovitis, acne,  
CC pustulosis, hyperostosis and osteomyelitis), uveitis and endophthalmitis.  
CC *P. acnes* is also involved in infections of bone, joints and the central  
CC nervous system, however it is particularly involved in the inflammatory  
CC lesions associated with acne vulgaris. A method for detecting the  
CC presence or absence of *P. acnes* in a patient comprises contacting a  
CC sample with a binding agent that binds to the proteins of the invention  
CC and determining the amount of bound protein in the sample. The  
CC polypeptides may be used as antigens in the production of antibodies  
CC specific for *P. acnes* proteins. These antibodies can be used to  
CC downregulate expression and activity of *P. acnes* polypeptides and  
CC therefore treat *P. acnes* infections. The antibodies may also be used as  
CC diagnostic agents for determining *P. acnes* presence, for example, by  
CC enzyme linked immunosorbent assay (ELISA). Note: The sequence data for  
CC this patent did not form part of the printed specification, but was  
CC obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 74 AA;  
Query Match 88.9%; Score 32; DB 4; Length 74;  
Best Local Similarity 100.0%; Pred. No. 7.6e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 AGGPC 5  
DB 29 AGGPC 33  
RESULT 37  
ABG29902 standard; protein; 74 AA.  
XX  
AC ABG29902;  
XX  
DT 18-FEB-2002 (first entry)  
XX  
DE Novel human diagnostic protein #29893.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PE 30-MAR-2001; 2001WO-US008631.  
XX  
PR 31-MAR-2000; 2000US-00540217.  
PR 23-AUG-2000; 2000US-00649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI; 2001-639362/73.  
DR N-PSDB; AAS94089.  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.  
XX  
PS Claim 20; SEQ ID NO 60261; 103pp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and polypeptide (II)

CC sequences. (I) is useful as hybridisation probes, polymerase chain  
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
CC and in recombinant production of (II). The polynucleotides are also used  
CC in diagnostics as expressed sequence tags for identifying expressed  
CC genes. (I) is useful in gene therapy techniques to restore normal  
CC activity of (II) or to treat disease states involving (II). (II) is  
CC useful for generating antibodies against it, detecting or quantitating a  
CC polypeptide in tissue, as molecular weight markers and as a food  
CC supplement. (II) and its binding partners are useful in medical imaging  
CC of sites expressing (II). (I) and (II) are useful for treating disorders  
CC involving aberrant protein expression or biological activity. The  
CC polypeptide and polynucleotide sequences have applications in  
CC diagnostic, forensics, gene mapping, identification of mutations  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic  
CC amino acid sequences of the invention. Note: The sequence data for this  
CC patent did not appear in the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 74 AA;  
Query Match 88.9%; Score 32; DB 4; Length 74;  
Best Local Similarity 100.0%; Pred. No. 7.6e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 AGGPC 5  
DB 39 AGGPC 43  
RESULT 38  
ABM61706  
ID ABM61706 standard; protein; 74 AA.  
XX  
AC ABM61706;  
XX  
DT 20-OCT-2003 (first entry)  
XX  
DE *Propionibacterium acnes* predicted ORF-encoded polypeptide #26382.  
XX  
KW Acne vulgaris; antiseborrheic; dermatological; antibacterial;  
KW immunostimulant; immune response; vaccine.  
XX  
OS *Propionibacterium acnes*.  
XX  
PN WO2003033515-A1.  
XX  
PD 24-APR-2003.  
XX  
PE 11-OCT-2002; 2002WO-US032727.  
XX  
PR 15-OCT-2001; 2001US-00978825.  
XX  
PA (CORI-) CORIXA CORP.  
XX  
PI Mitcham JL, Skeiky YAW, Persing DH, Bhatia A, Maisonneuve JL;  
PI Zhang Y, Wang S, Jen S, Lodes MJ, Benson DR, Jones R, Carter D;  
PI Barth B, Vallieve-Douglases J;  
XX  
DR WPI; 2003-381789/36.  
DR N-PSDB; ACF64590.  
XX  
XX New *Propionibacterium acnes* polypeptides and polynucleotides encoding the  
PT polypeptide, useful for diagnosing, preventing or treating acne vulgaris,  
PT or for stimulating an immune response specific for a *P. acnes* protein.  
XX  
PS Example 1, SEQ ID NO 26382, 1481pp; English.  
XX  
CC The invention relates to an isolated polynucleotide (ACF64435-ACF64733)  
CC encoding a *Propionibacterium acnes* protein. The invention also relates to  
CC polypeptides encoded by the polynucleotides (ABM55624-ABM64536) and to



CC immunogenic fragments of *P. acnes* polypeptides. The invention  
 CC additionally encompasses expression vectors and host cells comprising a  
 CC polynucleotide of the invention; antibodies against polypeptides of the  
 CC invention; fusion proteins comprising a polypeptide of the invention; a  
 CC method for stimulating an immune response specific for a *P. acnes*  
 CC polypeptide and an isolated T cell population comprising T cells prepared  
 CC via this method; a vaccine composition (comprising *P. acnes* polypeptides,  
 CC polynucleotides, antibodies, fusion proteins, T cell populations, or  
 CC antigen-presenting cells that express the polypeptide); a method and kit  
 CC for detecting or determining the presence or absence of *P. acnes* in a  
 CC patient; and a method for inhibiting the development of *P. acnes* in a  
 CC patient. The *P. acnes* polypeptides, polynucleotides, antibodies, fusion  
 CC proteins, T cell populations or antigen-presenting cells that express the  
 CC polypeptides are useful for diagnosing, preventing or treating *acne*  
 CC vulgaris, or for stimulating an immune response specific for a *P. acnes*  
 CC protein. The polynucleotides can also be used as probes or primers for  
 CC nucleic acid hybridization. The vaccine composition is useful for the  
 CC stimulation of an immune response against *P. acnes*, or for treating *acne*,  
 CC and the kit is useful for performing a diagnostic assay. The present  
 CC sequence represents a polypeptide predicted to be encoded by an ORF (open  
 CC reading frame) contained within the *P. acnes* polynucleotides of the  
 CC invention. Note: The sequence data for this patent did not form part of  
 CC the printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

SO Sequence 74 AA:

Query Match 88.9%; Score 32; DB 6; Length 74;  
 Best Local Similarity 100.0%; Pred. No. 7.6e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGPC 5  
 |||||  
 Db 29 AGGPC 33

RESULT 39  
 ADA57035  
 ID ADA57035 standard; protein; 88 AA.  
 AC ADA57035;  
 XX 20-NOV-2003 (first entry)  
 DT 20-NOV-2003 (first entry)  
 DE Human secreted protein #318.  
 XX  
 XX Immunosuppressive; antiinflammatory; antiaesthetic; antiallergic;  
 KM cytostatic; cerebroprotective; neuroprotective; nootropic;  
 KM cardiovascular; antiarteriosclerotic; gene therapy;  
 KM human secreted protein; immune disorder; inflammation;  
 KM respiratory disorder; cancer; CNS disorder; neurodegenerative disorders;  
 KM inflammatory bowel disease; nephritis; Crohn's disease; asthma; allergy;  
 KM multiple sclerosis; ischaemic brain injury; Parkinson's disease;  
 KM Alzheimer's disease; atherosclerosis; myocarditis; chromosome mapping;  
 KM tripe helix formation; antisense gene therapy; forensic biology.

OS Homo sapiens.  
 XX  
 XX WO2002102994-A2.  
 XX  
 XX 27-DEC-2002.  
 XX  
 XX 19-MAR-2002; 2002WC-US008278.  
 XX  
 XX 21-MAR-2001; 2001US-0277340P.  
 PR 19-JUL-2001; 2001US-0306171P.  
 PR 13-NOV-2001; 2001US-0331287P.  
 XX  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 PA Rosen CA, Ruben SM,  
 XX  
 XX WPI, 2003-167512/16.

DR N-PSDB; ADA56139.  
 XX  
 XX New human secreted polypeptides and polynucleotides, useful for  
 FT diagnosing, treating or preventing e.g. immune disorders, inflammatory  
 PT neurodegenerative disorders, cancers, CNS disorders, or  
 PT neurodegenerative disorders.

XX Claim 13; SEQ ID NO 1225; 1754pp; English.

XX The invention relates to 592 new human secreted polypeptides useful for  
 CC diagnosing, treating or preventing e.g. immune disorders, inflammatory  
 CC conditions, respiratory disorders, cancers, CNS disorders, or  
 CC neurodegenerative disorders, or polypeptides comprising an amino acid  
 CC sequence at least 95% identical to the new sequences. The polypeptides,  
 CC antibodies or antibody fragments that bind to the polypeptides, nucleic  
 CC acids encoding the polypeptides, agonists or antagonists that binds to  
 CC the polypeptide, are useful in preparing diagnostic or pharmaceutical  
 CC compositions for diagnosing, treating or preventing an e.g. immune  
 CC disorders, inflammatory conditions (e.g. inflammatory bowel disease,  
 CC nephritis or Crohn's disease), respiratory disorders (e.g. asthma and  
 CC allergy), cancers (e.g. gastric, ovarian or lung cancer), CNS disorders  
 CC (e.g. multiple sclerosis or ischaemic brain injury), neurodegenerative  
 CC disorders (e.g. Parkinson's disease or Alzheimer's disease), and  
 CC cardiovascular disorders (e.g. atherosclerosis or myocarditis). The  
 CC polynucleotides are useful for chromosome identification, chromosome  
 CC mapping, for controlling gene expression through triple helix formation  
 CC or antisense DNA or RNA, in gene therapy, for identifying individuals  
 CC from minute biological samples, in forensic biology, and as hybridization  
 CC probes. The polypeptides are useful for as molecular weight markers on  
 CC sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE)  
 CC gels, to raise antibodies, for testing biological activities, and for  
 CC treating or preventing neural disorders, immune system disorders,  
 CC muscular, reproductive, gastrointestinal, pulmonary, cardiovascular,  
 CC renal, proliferative and/or cancerous diseases. This sequence corresponds  
 CC to one of the polypeptide of the invention. Note: The sequence data for  
 CC this patent did form part of the printed specification, but was obtained  
 CC in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.

XX  
 XX Sequence 88 AA:

Query Match 88.9%; Score 32; DB 6; Length 88;  
 Best Local Similarity 100.0%; Pred. No. 8.8e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGPC 5  
 |||||  
 Db 58 AGGPC 62

RESULT 40  
 ADA40890  
 ID ADA40890 standard; protein; 88 AA.  
 AC ADA40890;  
 XX 20-NOV-2003 (first entry)  
 DT 20-NOV-2003 (first entry)  
 DE Human secreted protein.  
 XX  
 XX Human, secreted protein; cancer; hyperproliferative disorder;  
 KM rheumatoid arthritis; autoimmune disorder; haematopoietic disorder;  
 KM anaemia; allergic reaction; asthma; cardiovascular disorder;  
 KM wound healing; cytostatic; immunosuppressive; nootropic; neuroprotective;  
 KM antiviral; antiinflammatory; hepatotropic; antidiabetic; antiinflammatory;  
 KM vulnery; cardiant; gene therapy.

XX  
 XX Homo sapiens.  
 OS  
 XX WO2002102993-A2.  
 XX  
 XX 27-DEC-2002.

PF 19-MAR-2002; 2002WO-US008123.  
 XX  
 PR 21-MAR-2001; 2001US-0277340P.  
 PR 19-JUL-2001; 2001US-0306171P.  
 PR 13-NOV-2001; 2001US-0331287P.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Rosen CA, Ruben SM;  
 XX  
 DR WPI; 2003-175238/17.  
 XX  
 PT New human secreted proteins and nucleic acid molecules, useful for  
 PT preparing a diagnostic or pharmaceutical composition for diagnosing,  
 PT preventing or treating cancer or other hyperproliferative disorder,  
 PT asthma, allergies or AIDS.  
 XX  
 PS Claim 1; SEQ ID NO 1272; 3205bp; English.  
 XX  
 CC The invention relates to novel genes ADA39629-ADA40565 and proteins  
 CC ADA40566-ADA41501 for human secreted proteins, useful for preventing,  
 CC treating or ameliorating medical conditions e.g. by protein or gene  
 CC therapy. The polypeptides, nucleic acid molecules, antibodies or their  
 CC fragments, and agonists or antagonists that bind to the polypeptide are  
 CC useful for preparing a diagnostic or pharmaceutical composition for  
 CC diagnosing or treating cancer or other hyperproliferative disorder. The  
 CC polypeptides and nucleic acid molecules are also useful for detecting,  
 CC preventing, diagnosing, prognosticating, treating or ameliorating cancer  
 CC or other hyperproliferative disorders including neoplasms, autoimmune  
 CC disorders (e.g. diabetes, rheumatoid arthritis, systemic lupus  
 CC erythematosus, multiple sclerosis, autoimmune thyroiditis or haemolytic  
 CC anaemia), haematopoietic or haematological disorders (e.g. anaemia,  
 CC thrombocytopenia), allergic reactions including asthma or eczema,  
 CC inflammatory disorders (e.g. ischaemia-reperfusion injury, inflammatory  
 CC bowel disease or Crohn's disease), neurodegenerative disorders (e.g.  
 CC Alzheimer's disease or Parkinson's disease), cardiovascular disorders  
 CC (e.g. atherosclerosis, myocarditis), infectious diseases (bacterial,  
 CC fungal or viral infections including HIV/AIDS), or wound healing and  
 CC disorders of epithelial cell proliferation. The nucleic acids are also  
 CC useful for chromosome identification, radiation hybrid mapping or long-  
 CC range restriction mapping, as molecular weight markers, or as  
 CC hybridization or diagnostic probes. The polypeptides and antibodies are  
 CC useful for providing immunological probes for differential identification  
 CC of the tissues immunohistochemistry assays. Note: The sequence data for  
 CC this patent did not form part of the printed specification, but was  
 CC obtained in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences.  
 XX  
 SQ Sequence 88 AA;

Query Match 88.9%; Score 32; DB 6; Length 88;  
 Best Local Similarity 100.0%; Pred. No. 8.8e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGGPC 5  
 |||||  
 Db 58 AGGPC 62

Search completed: December 30, 2004, 15:08:57  
 Job time : 10.7879 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: December 30, 2004, 15:15:03 ; Search time 156 Seconds

(without alignments)  
32.194 Million cell updates/sec

Title: US-10-719-385-20

Perfect score: 14

Sequence: 1 MIRSKITSVLSFC 14

Scoring table: Gapop 60.0 , Gapext 60.0

Searched: 2002273 seqs, 358729299 residues

Word size : 0

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 150 summaries

Database : A\_Geneseq\_23Sep04:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Match	Query Length	DB ID	Description
1	14	100.0	14	8	ADO71787 BFLP0169
2	14	100.0	1752	8	ADO71766 BFLP0169
3	14	100.0	1753	8	ADO71777 BFLP0169
4	14	100.0	1753	8	ADO71767 BFLP0169
5	14	100.0	1753	8	ADO71775 BFLP0169
6	14	100.0	1753	8	ADO71768 BFLP0169
7	14	100.0	1753	8	ADO71771 BFLP0169
8	14	100.0	1753	8	ADO71780 BFLP0169
9	14	100.0	1753	8	ADO71772 BFLP0169
10	14	100.0	1753	8	ADO71764 BFLP0169
11	14	100.0	1753	8	ADO71765 BFLP0169
12	14	100.0	1753	8	ADO71773 BFLP0169
13	14	100.0	1753	8	ADO71778 BFLP0169
14	14	100.0	1753	8	ADO71769 BFLP0169
15	14	100.0	1753	8	ADO71770 BFLP0169
16	14	100.0	1753	8	ADO71774 BFLP0169
17	14	100.0	1753	8	ADO71779 BFLP0169
18	14	100.0	1753	8	ADO71763 BFLP0169
19	14	100.0	1753	8	ADO71776 BFLP0169
20	14	100.0	1753	8	ADO71777 BFLP0169
21	14	100.0	1753	8	ADO71769 BFLP0169
22	14	100.0	1753	8	ADO71770 BFLP0169
23	14	100.0	1753	8	ADO71774 BFLP0169
24	14	100.0	1753	8	ADO71779 BFLP0169
25	14	100.0	1753	8	ADO71763 BFLP0169
26	14	100.0	1753	8	ADO71776 BFLP0169
27	14	100.0	1753	8	ADO71777 BFLP0169
28	14	100.0	1753	8	ADO71769 BFLP0169
29	14	100.0	1753	8	ADO71770 BFLP0169
30	14	100.0	1753	8	ADO71774 BFLP0169
31	14	100.0	1753	8	ADO71779 BFLP0169
32	14	100.0	1753	8	ADO71763 BFLP0169
33	14	100.0	1753	8	ADO71776 BFLP0169
34	14	100.0	1753	8	ADO71777 BFLP0169
35	14	100.0	1753	8	ADO71769 BFLP0169
36	14	100.0	1753	8	ADO71770 BFLP0169
37	14	100.0	1753	8	ADO71774 BFLP0169
38	14	100.0	1753	8	ADO71779 BFLP0169
39	14	100.0	1753	8	ADO71763 BFLP0169
40	14	100.0	1753	8	ADO71776 BFLP0169
41	14	100.0	1753	8	ADO71777 BFLP0169
42	14	100.0	1753	8	ADO71769 BFLP0169
43	14	100.0	1753	8	ADO71770 BFLP0169
44	14	100.0	1753	8	ADO71774 BFLP0169
45	14	100.0	1753	8	ADO71779 BFLP0169
46	14	100.0	1753	8	ADO71763 BFLP0169
47	14	100.0	1753	8	ADO71776 BFLP0169
48	14	100.0	1753	8	ADO71777 BFLP0169
49	14	100.0	1753	8	ADO71769 BFLP0169
50	14	100.0	1753	8	ADO71770 BFLP0169
51	14	100.0	1753	8	ADO71774 BFLP0169
52	14	100.0	1753	8	ADO71779 BFLP0169
53	14	100.0	1753	8	ADO71763 BFLP0169
54	14	100.0	1753	8	ADO71776 BFLP0169
55	14	100.0	1753	8	ADO71777 BFLP0169
56	14	100.0	1753	8	ADO71769 BFLP0169
57	14	100.0	1753	8	ADO71770 BFLP0169
58	14	100.0	1753	8	ADO71774 BFLP0169
59	14	100.0	1753	8	ADO71779 BFLP0169
60	14	100.0	1753	8	ADO71763 BFLP0169
61	14	100.0	1753	8	ADO71776 BFLP0169
62	14	100.0	1753	8	ADO71777 BFLP0169
63	14	100.0	1753	8	ADO71769 BFLP0169
64	14	100.0	1753	8	ADO71770 BFLP0169
65	14	100.0	1753	8	ADO71774 BFLP0169
66	14	100.0	1753	8	ADO71779 BFLP0169
67	14	100.0	1753	8	ADO71763 BFLP0169
68	14	100.0	1753	8	ADO71776 BFLP0169
69	14	100.0	1753	8	ADO71777 BFLP0169
70	14	100.0	1753	8	ADO71769 BFLP0169
71	14	100.0	1753	8	ADO71770 BFLP0169
72	14	100.0	1753	8	ADO71774 BFLP0169
73	14	100.0	1753	8	ADO71779 BFLP0169
74	14	100.0	1753	8	ADO71763 BFLP0169
75	14	100.0	1753	8	ADO71776 BFLP0169
76	14	100.0	1753	8	ADO71777 BFLP0169
77	14	100.0	1753	8	ADO71769 BFLP0169
78	14	100.0	1753	8	ADO71770 BFLP0169
79	14	100.0	1753	8	ADO71774 BFLP0169
80	14	100.0	1753	8	ADO71779 BFLP0169
81	14	100.0	1753	8	ADO71763 BFLP0169
82	14	100.0	1753	8	ADO71776 BFLP0169
83	14	100.0	1753	8	ADO71777 BFLP0169
84	14	100.0	1753	8	ADO71769 BFLP0169
85	14	100.0	1753	8	ADO71770 BFLP0169
86	14	100.0	1753	8	ADO71774 BFLP0169
87	14	100.0	1753	8	ADO71779 BFLP0169
88	14	100.0	1753	8	ADO71763 BFLP0169
89	14	100.0	1753	8	ADO71776 BFLP0169
90	14	100.0	1753	8	ADO71777 BFLP0169
91	14	100.0	1753	8	ADO71769 BFLP0169
92	14	100.0	1753	8	ADO71770 BFLP0169
93	14	100.0	1753	8	ADO71774 BFLP0169
94	14	100.0	1753	8	ADO71779 BFLP0169
95	14	100.0	1753	8	ADO71763 BFLP0169
96	14	100.0	1753	8	ADO71776 BFLP0169
97	14	100.0	1753	8	ADO71777 BFLP0169
98	14	100.0	1753	8	ADO71769 BFLP0169
99	14	100.0	1753	8	ADO71770 BFLP0169
100	14	100.0	1753	8	ADO71774 BFLP0169

26	7	50.0	431	8	ADO25474	Ado25474 E. faecali
27	7	50.0	1019	4	AA922745	AA922745 C. glutami
28	7	50.0	1019	5	AB879444	AB879444 Coryneb
29	7	50.0	1019	7	ADL65625	ADL65625 C. glutam
30	6	42.9	12	4	AA559995	AA559995 Internali
31	6	42.9	12	7	ADA88830	ADA88830 Internali
32	6	42.9	17	5	ABG78459	ABG78459 HIRRS11
33	6	42.9	77	3	ABG00649	ABG00649 Human sec
34	6	42.9	85	4	ABG06877	ABG06877 Novel hum
35	6	42.9	90	4	AA939937	AA939937 Propionib
36	6	42.9	90	6	ABM36456	ABM36456 Propionib
37	6	42.9	130	6	ABU17550	ABU17550 Protein e
38	6	42.9	133	7	ADC94908	ADC94908 E. faecali
39	6	42.9	135	7	ADF04124	ADF04124 Bacterial
40	6	42.9	139	5	ABP10237	ABP10237 Human ORP
41	6	42.9	143	7	ADC88364	ADC88364 Ribosomal
42	6	42.9	183	7	ADP6790	ADP6790 Bacterial
43	6	42.9	184	8	ADO03923	ADO03923 Rabbit fu
44	6	42.9	186	8	ADO03922	ADO03922 Rat full
45	6	42.9	192	5	ADK34600	ADK34600 Novel hum
46	6	42.9	204	4	AA93258	AA93258 Human pol
47	6	42.9	204	4	ADL30674	ADL30674 Human pro
48	6	42.9	207	4	AAU19352	AAU19352 Human G p
49	6	42.9	216	5	AA61671	AA61671 Arabidops
50	6	42.9	218	5	ABP27242	ABP27242 Streptoco
51	6	42.9	238	8	ADK47075	ADK47075 Streptoco
52	6	42.9	242	3	AA628834	AA628834 Arabidops
53	6	42.9	242	3	AA635588	AA635588 Arabidops
54	6	42.9	270	7	ADC94107	ADC94107 E. faecali
55	6	42.9	284	3	AA606151	AA606151 Arabidops
56	6	42.9	289	3	AA627262	AA627262 Arabidops
57	6	42.9	289	3	AA640485	AA640485 Arabidops
58	6	42.9	289	5	AA625109	AA625109 Arabidops
59	6	42.9	289	5	AAU72581	AAU72581 Arabidops
60	6	42.9	289	8	ADN72347	ADN72347 Thale cre
61	6	42.9	298	8	ADN61913	ADN61913 Human nov
62	6	42.9	302	4	AA604555	AA604555 Human G-p
63	6	42.9	308	6	ADK50458	ADK50458 Human pro
64	6	42.9	311	4	AA672083	AA672083 Human oif
65	6	42.9	311	4	AAU24523	AAU24523 Human oif
66	6	42.9	311	5	AAU58562	AAU58562 Human G p
67	6	42.9	311	5	AAU58563	AAU58563 Human G p
68	6	42.9	311	5	ABP95796	ABP95796 Human GPC
69	6	42.9	311	5	AAU85147	AAU85147 G-coupled
70	6	42.9	314	5	ABP51568	ABP51568 Human G p
71	6	42.9	314	7	ADL04261	ADL04261 Human G-p
72	6	42.9	316	6	ABU19105	ABU19105 Pathogen
73	6	42.9	317	5	ADK36453	ADK36453 Human PAM
74	6	42.9	317	8	ADU19375	ADU19375 Human PAM
75	6	42.9	321	5	AB890173	AB890173 Human pol
76	6	42.9	321	5	ADU19416	ADU19416 Human PAM
77	6	42.9	321	8	ADU19416	ADU19416 Human PAM
78	6	42.9	322	5	ABP28189	ABP28189 Streptoco
79	6	42.9	326	3	AA606150	AA606150 Arabidops
80	6	42.9	326	6	ABR56928	ABR56928 Arabidops
81	6	42.9	326	8	AAU98390	AAU98390 Arabidops
82	6	42.9	327	3	AA827261	AA827261 Arabidops
83	6	42.9	329	4	AAU89960	AAU89960 Human imm
84	6	42.9	332	6	ABU18942	ABU18942 Pathogen
85	6	42.9	332	6	ABW72372	ABW72372 Staphyloc
86	6	42.9	336	3	AA606149	AA606149 Arabidops
87	6	42.9	340	5	ABP99397	ABP99397 Arabidops
88	6	42.9	344	7	ADC95919	ADC95919 E. faecali
89	6	42.9	360	3	AAV77470	AAV77470 Human deu
90	6	42.9	360	5	ABU61835	ABU61835 Human de-
91	6	42.9	374	5	AAU80496	AAU80496 Human G-C
92	6	42.9	376	5	ABP30123	ABP30123 Streptoco
93	6	42.9	384	7	ADJ08050	ADJ08050 Novel pro
94	6	42.9	387	7	ADJ70426	ADJ70426 Human hea
95	6	42.9	403	5	ABG31768	ABG31768 Nuclear h
96	6	42.9	404	6	ABU02498	ABU02498 S. pneumo
97	6	42.9	404	6	ABP81646	ABP81646 Streptoco
98	6	42.9	413	8	ADL66350	ADL66350 Human col

99	6	42.9	413	8	ADN9857	Adn9857 Novel hum
100	6	42.9	416	5	ABP27243	Abp27243 Streptoco
101	6	42.9	429	3	AGS3587	AgS3587 Arabidops
102	6	42.9	429	3	AGS28833	AgS28833 Arabidops
103	6	42.9	430	6	ABP75913	Abp75913 Human sec
104	6	42.9	431	6	ABU29849	Abu29849 Protein e
105	6	42.9	433	3	AGS28832	AgS28832 Arabidops
106	6	42.9	433	3	AGS3586	AgS3586 Arabidops
107	6	42.9	438	8	ADN9854	Adn9854 Novel hum
108	6	42.9	447	8	ADP49384	Adp49384 Xanthomon
109	6	42.9	468	7	ADMT2042	Adm2042 Hyperther
110	6	42.9	482	3	AGS31506	AgS31506 Arabidops
111	6	42.9	482	5	ABN90925	Abn90925 Herbi-cida
112	6	42.9	494	2	AMP30711	AmP30711 Human ubi
113	6	42.9	496	8	ADP47895	Adp47895 Human CAT
114	6	42.9	497	7	ADC31203	AdC31203 Human nov
115	6	42.9	497	7	ADC27031	AdC27031 Human deu
116	6	42.9	501	3	AGA47662	AgA47662 Arabidops
117	6	42.9	519	4	ABG12217	Abg12217 Novel hum
118	6	42.9	519	4	ABG05882	Abg05882 Novel hum
119	6	42.9	527	3	AGS35833	AgS35833 Arabidops
120	6	42.9	530	3	AAV77471	AaV77471 Human deu
121	6	42.9	530	3	AAAG64049	AaG64049 Human deu
122	6	42.9	530	6	ABU61836	Abu61836 Updated h
123	6	42.9	530	6	ABU10207	Abu10207 Human ubi
124	6	42.9	530	6	ABR43250	AbR43250 Human PMW
125	6	42.9	530	7	ADC27021	AdC27021 Human deu
126	6	42.9	530	7	ADC27033	AdC27033 Human deu
127	6	42.9	530	7	ADC27019	AdC27019 Human deu
128	6	42.9	530	7	ADC27037	AdC27037 Human deu
129	6	42.9	530	7	ADC27039	AdC27039 Human deu
130	6	42.9	530	7	ADC27009	AdC27009 Human deu
131	6	42.9	530	7	ADC27035	AdC27035 Human deu
132	6	42.9	530	7	ADC27015	AdC27015 Human deu
133	6	42.9	530	8	ADN9856	Adn9856 Novel hum
134	6	42.9	538	4	ABG07586	AbG07586 Novel hum
135	6	42.9	539	4	AAAB31604	AaB31604 Amino act
136	6	42.9	540	2	AAW07871	AaW07871 GDU (or G
137	6	42.9	558	6	ABU25727	AbU25727 Protein e
138	6	42.9	565	7	ADC27023	AdC27023 Human deu
139	6	42.9	566	6	ABR53129	AbR53129 Protein s
140	6	42.9	566	7	ADK62560	AdK62560 Disease c
141	6	42.9	574	7	ADC27017	AdC27017 Human deu
142	6	42.9	578	2	AAV09065	AaV09065 Human com
143	6	42.9	584	8	ADH17324	AdH17324 Human NOV
144	6	42.9	587	6	ABOS8812	AbOS8812 Human gen
145	6	42.9	587	6	ABR53423	AbR53423 Protein s
146	6	42.9	587	4	ADK64684	AdK64684 Disease c
147	6	42.9	597	4	ABR11735	AbB11735 Human vas
148	6	42.9	597	7	ADBO3343	AdBO3343 Novel pro
149	6	42.9	597	8	ADM87677	AdM87677 Human EST
150	6	42.9	615	3	AAAG35832	AaG35832 Arabidops

## ALIGNMENTS

## RESULT 1

ADO71787 standard; protein; 14 AA.

ADO71787

26-AUG-2004 (first entry)

BFLP0169 protein fragment, SEQ ID 20.

Antiinflammatory; Dermatological; Immunosuppressive; human; BFLP0169;

Kidney; lupus nephritis; autoimmune disorder; lupus.

Homo sapiens.

WO2004048521-A2.

XX 10-JUN-2004.

PD 21-NOV-2003; 2003WO-US037339.

XX 21-NOV-2002; 2002US-0428094P.

XX (AMHP) MYETH.

XX O'Loole M, Mounts WM, Shojaae N;

XX WPI; 2004-441163/41.

PT New isolated BFLP0169 nucleic acids and proteins, useful for diagnosing

PT or treating autoimmune disorder, e.g. lupus or lupus nephritis.

PS Claim 20; Page; 83pp; English.

XX The present invention relates to human BFLP0169 (ADO71763) and its coding

CC sequence (ADO71762). BFLP0169 gene expression is increased in kidney

CC tissue in mice with lupus nephritis. The BFLP0169 sequences are useful

CC for diagnosing and treating autoimmune disorder, preferably lupus or

CC lupus nephritis. They are also useful in the generation of antibodies

CC that bind immunospecifically to the nucleic acids. The present sequence

CC was used to illustrate the invention. Note: The present sequence was not

CC shown in the specification, but was derived from information given on

CC page 13 and ADO71763.

SQ Sequence 14 AA;

Query Match 100.0%; Score 14; DB 8; Length 14;

Best Local Similarity 100.0%; Pred. No. 6; 1e-08;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MIRSKITSVLSFC 14

Db 1 MIRSKITSVLSFC 14

## RESULT 2

ADO71766 standard; protein; 1752 AA.

ADO71766

26-AUG-2004 (first entry)

BFLP0169 variant R267K, SEQ ID 5.

Antiinflammatory; Dermatological; Immunosuppressive; human; BFLP0169;

Kidney; lupus nephritis; autoimmune disorder; lupus; mutant; mutacin.

Homo sapiens.

Synthetic.

Location/Qualifiers

Key MISC-difference 267 /note= "Wild-type R replaced with K"

WO2004048521-A2.

10-JUN-2004.

21-NOV-2003; 2003WO-US037339.

21-NOV-2002; 2002US-0428094P.

(AMHP) MYETH.

O'Loole M, Mounts WM, Shojaae N;

WPI; 2004-441163/41.

PT New isolated BFLP0169 nucleic acids and proteins, useful for diagnosing  
PT or treating autoimmune disorder, e.g. lupus or lupus nephritis.  
XX  
PS Example 4; Page 64-65; 83pp; English.

CC The present invention relates to human BFLP0169 (ADO71763) and its coding  
CC sequence (ADO71762). BFLP0169 gene expression is increased in kidney  
CC tissue in mice with lupus nephritis. The BFLP0169 sequences are useful  
CC for diagnosing and treating autoimmune disorder, preferably lupus or  
CC lupus nephritis. They are also useful in the generation of antibodies  
CC that bind immunospecifically to the nucleic acids. The present sequence  
CC is a BFLP0169 variant.  
XX  
SQ Sequence 1752 AA;

Query Match 100.0%; Score 14; DB 8; Length 1752;  
Best Local Similarity 100.0%; Pred. No. 3.9e-06;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MIRSKITSVLSFC 14  
| | | | | | | | | | | | | | | | | |  
Db 1 MIRSKITSVLSFC 14

RESULT 3  
ADO71777  
ID ADO71777 standard; protein; 1753 AA.  
XX  
AC ADO71777;  
XX  
DT 26-AUG-2004 (first entry)  
XX  
DE BFLP0169 variant K1331R, SEQ ID 16.  
XX  
KM Antinflammatory; Dermatological; Immunosuppressive; human; BFLP0169;  
XX kidney; lupus nephritis; autoimmune disorder; lupus; mutant; mutein.  
XX  
OS Homo sapiens.  
OS Synthetic.  
OS  
FH Key Location/Qualifiers  
FH Misc-difference 1331  
FT /note= "Wild-type K replaced with R"

PN MO2004048521-A2.  
XX  
PD 10-JUN-2004.  
XX  
PF 21-NOV-2003; 2003WO-US037339.  
XX  
PR 21-NOV-2002; 2002US-0428094P.  
XX  
PA (AMHP ) WYETH.  
XX  
PI O'coole M, Mounts WM, Shojaae N;  
XX  
DR WPI; 2004-441163/41.  
XX  
PT New isolated BFLP0169 nucleic acids and proteins, useful for diagnosing  
PT or treating autoimmune disorder, e.g. lupus or lupus nephritis.  
XX  
PS Example 15; Page 71; 83pp; English.

CC The present invention relates to human BFLP0169 (ADO71763) and its coding  
CC sequence (ADO71762). BFLP0169 gene expression is increased in kidney  
CC tissue in mice with lupus nephritis. The BFLP0169 sequences are useful  
CC for diagnosing and treating autoimmune disorder, preferably lupus or  
CC lupus nephritis. They are also useful in the generation of antibodies  
CC that bind immunospecifically to the nucleic acids. The present sequence  
CC is a BFLP0169 variant.  
XX  
SQ Sequence 1753 AA;

Query Match 100.0%; Score 14; DB 8; Length 1753;  
Best Local Similarity 100.0%; Pred. No. 3.9e-06;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MIRSKITSVLSFC 14  
| | | | | | | | | | | | | | | | | |  
Db 1 MIRSKITSVLSFC 14

RESULT 4  
ADO71767  
ID ADO71767 standard; protein; 1753 AA.  
XX  
AC ADO71767;  
XX  
DT 26-AUG-2004 (first entry)  
XX  
DE BFLP0169 variant Q355N, SEQ ID 6.  
XX  
KM Antinflammatory; Dermatological; Immunosuppressive; human; BFLP0169;  
XX kidney; lupus nephritis; autoimmune disorder; lupus; mutant; mutein.  
XX  
OS Homo sapiens.  
OS Synthetic.  
OS  
FH Key Location/Qualifiers  
FH Misc-difference 355  
FT /note= "Wild-type Q replaced with N"

PN MO2004048521-A2.  
XX  
PD 10-JUN-2004.  
XX  
PF 21-NOV-2003; 2003WO-US037339.  
XX  
PR 21-NOV-2002; 2002US-0428094P.  
XX  
PA (AMHP ) WYETH.  
XX  
PI O'coole M, Mounts WM, Shojaae N;  
XX  
DR WPI; 2004-441163/41.  
XX  
PT New isolated BFLP0169 nucleic acids and proteins, useful for diagnosing  
PT or treating autoimmune disorder, e.g. lupus or lupus nephritis.  
XX  
PS Example 5; Page 65; 83pp; English.

CC The present invention relates to human BFLP0169 (ADO71763) and its coding  
CC sequence (ADO71762). BFLP0169 gene expression is increased in kidney  
CC tissue in mice with lupus nephritis. The BFLP0169 sequences are useful  
CC for diagnosing and treating autoimmune disorder, preferably lupus or  
CC lupus nephritis. They are also useful in the generation of antibodies  
CC that bind immunospecifically to the nucleic acids. The present sequence  
CC is a BFLP0169 variant.  
XX  
SQ Sequence 1753 AA;

Query Match 100.0%; Score 14; DB 8; Length 1753;  
Best Local Similarity 100.0%; Pred. No. 3.9e-06;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MIRSKITSVLSFC 14  
| | | | | | | | | | | | | | | | | |  
Db 1 MIRSKITSVLSFC 14

RESULT 5  
ADO71775  
ID ADO71775 standard; protein; 1753 AA.  
XX  
AC ADO71775;  
XX

DT 26-AUG-2004 (first entry)  
 XX  
 DE BFLP0169 variant A11528, SEQ ID 14.  
 XX  
 XX Antiinflammatory; Dermatological; Immunosuppressive; human; BFLP0169;  
 KM kidney; lupus nephritis; autoimmune disorder; lupus; mutant; mutein.  
 XX  
 OS Homo sapiens.  
 OS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 1152 /note= "Wild-type A replaced with S"  
 PT  
 PN WO2004048521-A2.  
 XX  
 PD 10-JUN-2004.  
 XX  
 PF 21-NOV-2003; 2003WO-US037339.  
 XX  
 PR 21-NOV-2002; 2002US-0428094P.  
 XX  
 PA (AMHP ) WYETH.  
 XX  
 PI O'coole M, Mounts WM, Shojaee N;  
 XX  
 DR WPI; 2004-441163/41.  
 XX  
 PT New isolated BFLP0169 nucleic acids and proteins, useful for diagnosing  
 or treating autoimmune disorder, e.g. lupus or lupus nephritis.  
 XX  
 PS Example 13; Page 70; 83pp; English.  
 XX  
 CC The present invention relates to human BFLP0169 (ADO71763) and its coding  
 CC sequence (ADO71762). BFLP0169 gene expression is increased in kidney  
 CC tissue in mice with lupus nephritis. The BFLP0169 sequences are useful  
 CC for diagnosing and treating autoimmune disorder, preferably lupus or  
 CC lupus nephritis. They are also useful in the generation of antibodies  
 CC that bind immunospecifically to the nucleic acids. The present sequence  
 CC is a BFLP0169 variant.  
 XX  
 SQ Sequence 1753 AA;  
 Query Match 100.0%; Score 14; DB 8; Length 1753;  
 Best Local Similarity 100.0%; Pred. No. 3.9e-06;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MIRSKITSVLSFC 14  
 DB 1 MIRSKITSVLSFC 14  
 RESULT 6  
 ADO71768  
 ID ADO71768 standard; protein; 1753 AA.  
 XX  
 AC ADO71768;  
 XX  
 DT 26-AUG-2004 (first entry)  
 XX  
 DE BFLP0169 variant H446R, SEQ ID 7.  
 XX  
 XX Antiinflammatory; Dermatological; Immunosuppressive; human; BFLP0169;  
 KM kidney; lupus nephritis; autoimmune disorder; lupus; mutant; mutein.  
 XX  
 OS Homo sapiens.  
 OS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 446 /note= "Wild-type H replaced with R"  
 PT  
 PN WO2004048521-A2.

XX  
 PD 10-JUN-2004.  
 XX  
 PF 21-NOV-2003; 2003WO-US037339.  
 XX  
 PR 21-NOV-2002; 2002US-0428094P.  
 XX  
 PA (AMHP ) WYETH.  
 XX  
 PI O'coole M, Mounts WM, Shojaee N;  
 XX  
 DR WPI; 2004-441163/41.  
 XX  
 PT New isolated BFLP0169 nucleic acids and proteins, useful for diagnosing  
 or treating autoimmune disorder, e.g. lupus or lupus nephritis.  
 XX  
 PS Example 6; Page 66; 83pp; English.  
 XX  
 CC The present invention relates to human BFLP0169 (ADO71763) and its coding  
 CC sequence (ADO71762). BFLP0169 gene expression is increased in kidney  
 CC tissue in mice with lupus nephritis. The BFLP0169 sequences are useful  
 CC for diagnosing and treating autoimmune disorder, preferably lupus or  
 CC lupus nephritis. They are also useful in the generation of antibodies  
 CC that bind immunospecifically to the nucleic acids. The present sequence  
 CC is a BFLP0169 variant.  
 XX  
 SQ Sequence 1753 AA;  
 Query Match 100.0%; Score 14; DB 8; Length 1753;  
 Best Local Similarity 100.0%; Pred. No. 3.9e-06;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MIRSKITSVLSFC 14  
 DB 1 MIRSKITSVLSFC 14  
 RESULT 7  
 ADO71771  
 ID ADO71771 standard; protein; 1753 AA.  
 XX  
 AC ADO71771;  
 XX  
 DT 26-AUG-2004 (first entry)  
 XX  
 DE BFLP0169 variant N749D, SEQ ID 10.  
 XX  
 XX Antiinflammatory; Dermatological; Immunosuppressive; human; BFLP0169;  
 KM kidney; lupus nephritis; autoimmune disorder; lupus; mutant; mutein.  
 XX  
 OS Homo sapiens.  
 OS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 749 /note= "Wild-type N replaced with D"  
 PT  
 PN WO2004048521-A2.  
 XX  
 PD 10-JUN-2004.  
 XX  
 PF 21-NOV-2003; 2003WO-US037339.  
 XX  
 PR 21-NOV-2002; 2002US-0428094P.  
 XX  
 PA (AMHP ) WYETH.  
 XX  
 PI O'coole M, Mounts WM, Shojaee N;  
 XX  
 DR WPI; 2004-441163/41.  
 XX  
 PT New isolated BFLP0169 nucleic acids and proteins, useful for diagnosing  
 or treating autoimmune disorder, e.g. lupus or lupus nephritis.

```

XX Example 9, Page 67-68; 83bp; English.
XX
XX
CC The present invention relates to human BFLP0169 (ADO71763) and its coding
CC sequence (ADO71762). BFLP0169 gene expression is increased in kidney
CC tissue in mice with lupus nephritis. The BFLP0169 sequences are useful
CC for diagnosing and treating autoimmune disorder, preferably lupus or
CC lupus nephritis. They are also useful in the generation of antibodies
CC that bind immunospecifically to the nucleic acids. The present sequence
CC is a BFLP0169 variant.
XX
XX Sequence 1753 AA;
SQ
Query Match 100.0%; Score 14; DB 8; Length 1753;
Best Local Similarity 100.0%; Pred. No. 3.9e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MIRSKITSVLSFC 14
1 MIRSKITSVLSFC 14
Db
RESULT 8
ADO71780 ID ADO71780 standard; protein; 1753 AA.
XX
XX ADO71780;
AC
XX 26-AUG-2004 (first entry)
DT
XX BFLP0169 variant F1706H, SEQ ID 19.
DE
XX
XX Antinflammatory; Dermatological; Immunosuppressive; human; BFLP0169;
KM kidney; lupus nephritis; autoimmune disorder; lupus; mutant; mutein.
XX
XX Homo sapiens.
OS
XX Synthetic.
OS
XX Key Location/Qualifiers
FH Misc-difference 1706 /note= "Wild-type F replaced with H"
FT
XX
XX MO2004048521-A2.
FN
XX
XX 10-JUN-2004.
PD
XX
XX 21-NOV-2003; 2003WO-US037339.
PE
XX
XX 21-NOV-2002; 2002US-0428094P.
PR
XX
XX (AMHP ) WYETH.
PA
XX
XX O'coole M, Mounts WM, Shojaaee N;
PI
XX
XX WPI; 2004-441163/41.
DR
XX
XX New isolated BFLP0169 nucleic acids and proteins, useful for diagnosing
PT or treating autoimmune disorder, e.g. lupus or lupus nephritis.
PT
XX
XX Example 18; Page 73; 83bp; English.
PS
XX
XX The present invention relates to human BFLP0169 (ADO71763) and its coding
CC sequence (ADO71762). BFLP0169 gene expression is increased in kidney
CC tissue in mice with lupus nephritis. The BFLP0169 sequences are useful
CC for diagnosing and treating autoimmune disorder, preferably lupus or
CC lupus nephritis. They are also useful in the generation of antibodies
CC that bind immunospecifically to the nucleic acids. The present sequence
CC is a BFLP0169 variant.
XX
XX Sequence 1753 AA;
SQ
Query Match 100.0%; Score 14; DB 8; Length 1753;
Best Local Similarity 100.0%; Pred. No. 3.9e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MIRSKITSVLSFC 14
1 MIRSKITSVLSFC 14
Db
RESULT 9
ADO71772 ID ADO71772 standard; protein; 1753 AA.
XX
XX ADO71772;
AC
XX
XX 26-AUG-2004 (first entry)
DT
XX
XX BFLP0169 variant V838M, SEQ ID 11.
DE
XX
XX Antinflammatory; Dermatological; Immunosuppressive; human; BFLP0169;
KM kidney; lupus nephritis; autoimmune disorder; lupus; mutant; mutein.
XX
XX Homo sapiens.
OS
XX Synthetic.
OS
XX Key Location/Qualifiers
FH Misc-difference 838 /note= "Wild-type V replaced with M"
FT
XX
XX MO2004048521-A2.
FN
XX
XX 10-JUN-2004.
PD
XX
XX 21-NOV-2003; 2003WO-US037339.
PE
XX
XX 21-NOV-2002; 2002US-0428094P.
PR
XX
XX (AMHP ) WYETH.
PA
XX
XX O'coole M, Mounts WM, Shojaaee N;
PI
XX
XX WPI; 2004-441163/41.
DR
XX
XX New isolated BFLP0169 nucleic acids and proteins, useful for diagnosing
PT or treating autoimmune disorder, e.g. lupus or lupus nephritis.
PT
XX
XX Example 10; Page 68; 83bp; English.
PS
XX
XX The present invention relates to human BFLP0169 (ADO71763) and its coding
CC sequence (ADO71762). BFLP0169 gene expression is increased in kidney
CC tissue in mice with lupus nephritis. The BFLP0169 sequences are useful
CC for diagnosing and treating autoimmune disorder, preferably lupus or
CC lupus nephritis. They are also useful in the generation of antibodies
CC that bind immunospecifically to the nucleic acids. The present sequence
CC is a BFLP0169 variant.
XX
XX Sequence 1753 AA;
SQ
Query Match 100.0%; Score 14; DB 8; Length 1753;
Best Local Similarity 100.0%; Pred. No. 3.9e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MIRSKITSVLSFC 14
1 MIRSKITSVLSFC 14
Db
RESULT 10
ADO71764 ID ADO71764 standard; protein; 1753 AA.
XX
XX ADO71764;
AC
XX
XX 26-AUG-2004 (first entry)
DT
XX

```

```

DE BFLP0169 variant L78V, SEQ ID 3.
XX
XX Antinflammatory; Dermatological; Immunosuppressive; human; BFLP0169;
KW kidney; lupus nephritis; autoimmune disorder; lupus; mutant; mutein.
XX
XX Homo sapiens.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Misc-difference 78 /note= "Wild-type L replaced with V"
XX
XX WO2004048521-A2.
XX
XX 10-JUN-2004.
XX
XX 21-NOV-2003; 2003WO-US037339.
XX
XX 21-NOV-2002; 2002US-0428094P.
XX
XX (AMHP ) WYETH.
XX
XX O'coole M, Mounts WM, Shojasee N;
XX
XX WPI; 2004-441163/41.
XX
XX New isolated BFLP0169 nucleic acids and proteins, useful for diagnosing
XX or treating autoimmune disorder, e.g. lupus or lupus nephritis.
XX
XX Example 2, Page 63-64, 83pp; English.
XX
XX The present invention relates to human BFLP0169 (AD071763) and its coding
XX sequence (AD071762). BFLP0169 gene expression is increased in kidney
XX tissue in mice with lupus nephritis. The BFLP0169 sequences are useful
XX for diagnosing and treating autoimmune disorder, preferably lupus or
XX lupus nephritis. They are also useful in the generation of antibodies
XX that bind immunospecifically to the nucleic acids. The present sequence
XX is a BFLP0169 variant.
XX
XX Sequence 1753 AA;
XX
XX Query Match 100.0%; Score 14; DB 8; Length 1753;
XX Best Local Similarity 100.0%; Pred. No. 3.9e-06;
XX Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 MIRSKITSVLSFC 14
XX 1 MIRSKITSVLSFC 14
XX DB
XX
XX RESULT 11
XX AD071765
XX ID AD071765 standard; protein; 1753 AA.
XX
XX AC ADO71765;
XX
XX 26-AUG-2004 (first entry)
XX
XX BFLP0169 variant L198I, SEQ ID 4.
XX
XX Antinflammatory; Dermatological; Immunosuppressive; human; BFLP0169;
KW kidney; lupus nephritis; autoimmune disorder; lupus; mutant; mutein.
XX
XX Homo sapiens.
OS Synthetic.
XX
XX Key Location/Qualifiers
XX FH Misc-difference 198 /note= "Wild-type L replaced with I"
XX FT
XX WO2004048521-A2.
XX
XX 10-JUN-2004.
XX

```

```

XX
XX 21-NOV-2003; 2003WO-US037339.
XX
XX 21-NOV-2002; 2002US-0428094P.
XX
XX (AMHP ) WYETH.
XX
XX O'coole M, Mounts WM, Shojasee N;
XX
XX WPI; 2004-441163/41.
XX
XX New isolated BFLP0169 nucleic acids and proteins, useful for diagnosing
XX or treating autoimmune disorder, e.g. lupus or lupus nephritis.
XX
XX Example 3, Page 64, 83pp; English.
XX
XX The present invention relates to human BFLP0169 (AD071763) and its coding
XX sequence (AD071762). BFLP0169 gene expression is increased in kidney
XX tissue in mice with lupus nephritis. The BFLP0169 sequences are useful
XX for diagnosing and treating autoimmune disorder, preferably lupus or
XX lupus nephritis. They are also useful in the generation of antibodies
XX that bind immunospecifically to the nucleic acids. The present sequence
XX is a BFLP0169 variant.
XX
XX Sequence 1753 AA;
XX
XX Query Match 100.0%; Score 14; DB 8; Length 1753;
XX Best Local Similarity 100.0%; Pred. No. 3.9e-06;
XX Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 MIRSKITSVLSFC 14
XX 1 MIRSKITSVLSFC 14
XX DB
XX
XX RESULT 12
XX AD071773
XX ID AD071773 standard; protein; 1753 AA.
XX
XX AC ADO71773;
XX
XX 26-AUG-2004 (first entry)
XX
XX BFLP0169 variant G958T, SEQ ID 12.
XX
XX Antinflammatory; Dermatological; Immunosuppressive; human; BFLP0169;
KW kidney; lupus nephritis; autoimmune disorder; lupus; mutant; mutein.
XX
XX Homo sapiens.
OS Synthetic.
XX
XX Key Location/Qualifiers
XX FH Misc-difference 958 /note= "Wild-type G replaced with T"
XX FT
XX WO2004048521-A2.
XX
XX 10-JUN-2004.
XX
XX 21-NOV-2003; 2003WO-US037339.
XX
XX 21-NOV-2002; 2002US-0428094P.
XX
XX (AMHP ) WYETH.
XX
XX O'coole M, Mounts WM, Shojasee N;
XX
XX WPI; 2004-441163/41.
XX
XX New isolated BFLP0169 nucleic acids and proteins, useful for diagnosing
XX or treating autoimmune disorder, e.g. lupus or lupus nephritis.
XX
XX Example 11, Page 69, 83pp; English.
XX

```



XX The present invention relates to human BFLP0169 (ADO71763) and its coding  
 CC sequence (ADO71762). BFLP0169 gene expression is increased in kidney  
 CC tissue in mice with lupus nephritis. The BFLP0169 sequences are useful  
 CC for diagnosing and treating autoimmune disorder, preferably lupus or  
 CC lupus nephritis. They are also useful in the generation of antibodies  
 CC that bind immunospecifically to the nucleic acids. The present sequence  
 CC is a BFLP0169 variant.

XX SQ Sequence 1753 AA;

Query Match 100.0%; Score 14; DB 8; Length 1753;

Best Local Similarity 100.0%; Pred. No. 3.9e-06;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MIRSKITSVLSFC 14  
 Db 1 MIRSKITSVLSFC 14

RESULT 13

ADO71778 ID ADO71778 standard; protein; 1753 AA.

XX ADO71778;

XX 26-AUG-2004 (first entry)

XX BFLP0169 variant C1449Y, SEQ ID 17.

XX Antinflammatory; Dermatological; Immunosuppressive; human; BFLP0169;  
 KM kidney; lupus nephritis; autoimmune disorder; lupus; mutant; mutein.

XX Homo sapiens.  
 OS Synthetic.

XX Key Location/Qualifiers

FT Misc-difference 1449 /note= "Wild-type C replaced with Y"

XX MO2004048521-A2.

XX 10-JUN-2004.

XX 21-NOV-2003; 2003WO-US037339.

XX 21-NOV-2002; 2002US-0428094P.

XX (AMHP ) WYETH.

XX O'coole M, Mounts WM, Shojaei N;

XX WPI; 2004-441163/41.

XX New isolated BFLP0169 nucleic acids and proteins, useful for diagnosing  
 PT or treating autoimmune disorder, e.g. lupus or lupus nephritis.

XX Example 16; Page 72; 83pp; English.

XX The present invention relates to human BFLP0169 (ADO71763) and its coding  
 CC sequence (ADO71762). BFLP0169 gene expression is increased in kidney  
 CC tissue in mice with lupus nephritis. The BFLP0169 sequences are useful  
 CC for diagnosing and treating autoimmune disorder, preferably lupus or  
 CC lupus nephritis. They are also useful in the generation of antibodies  
 CC that bind immunospecifically to the nucleic acids. The present sequence  
 CC is a BFLP0169 variant.

XX SQ Sequence 1753 AA;

Query Match 100.0%; Score 14; DB 8; Length 1753;

Best Local Similarity 100.0%; Pred. No. 3.9e-06;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MIRSKITSVLSFC 14  
 Db 1 MIRSKITSVLSFC 14

RESULT 14

ADO71769 ID ADO71769 standard; protein; 1753 AA.

XX ADO71769;

XX 26-AUG-2004 (first entry)

XX BFLP0169 variant L553I, SEQ ID 8.

XX Antinflammatory; Dermatological; Immunosuppressive; human; BFLP0169;  
 KM kidney; lupus nephritis; autoimmune disorder; lupus; mutant; mutein.

XX Homo sapiens.  
 OS Synthetic.

XX Key Location/Qualifiers

FT Misc-difference 553 /note= "Wild-type L replaced with I"

XX MO2004048521-A2.

XX 10-JUN-2004.

XX 21-NOV-2003; 2003WO-US037339.

XX 21-NOV-2002; 2002US-0428094P.

XX (AMHP ) WYETH.

XX O'coole M, Mounts WM, Shojaei N;

XX WPI; 2004-441163/41.

XX New isolated BFLP0169 nucleic acids and proteins, useful for diagnosing  
 PT or treating autoimmune disorder, e.g. lupus or lupus nephritis.

XX Example 7; Page 66-67; 83pp; English.

XX The present invention relates to human BFLP0169 (ADO71763) and its coding  
 CC sequence (ADO71762). BFLP0169 gene expression is increased in kidney  
 CC tissue in mice with lupus nephritis. The BFLP0169 sequences are useful  
 CC for diagnosing and treating autoimmune disorder, preferably lupus or  
 CC lupus nephritis. They are also useful in the generation of antibodies  
 CC that bind immunospecifically to the nucleic acids. The present sequence  
 CC is a BFLP0169 variant.

XX SQ Sequence 1753 AA;

Query Match 100.0%; Score 14; DB 8; Length 1753;

Best Local Similarity 100.0%; Pred. No. 3.9e-06;

Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MIRSKITSVLSFC 14  
 Db 1 MIRSKITSVLSFC 14

RESULT 15

ADO71770 ID ADO71770 standard; protein; 1753 AA.

XX ADO71770;

XX 26-AUG-2004 (first entry)

XX BFLP0169 variant W634F, SEQ ID 9.

KM Anticardiolipin; Dermatological; Immunosuppressive; human; BFLP0169;  
 KM kidney; lupus nephritis; autoimmune disorder; lupus; mutant; mutain.  
 XX  
 OS Homo sapiens.  
 OS Synthetic.  
 XX  
 XX Key Location/Qualifiers  
 FT Misc-difference 634 /note= "Wild-type W replaced with F"  
 PT  
 XX  
 PN WO2004048521-A2.  
 XX  
 PD 10-JUN-2004.  
 XX  
 PF 21-NOV-2003; 2003WO-US037339.  
 XX  
 PR 21-NOV-2002; 2002US-0428094P.  
 XX  
 PA (AMHP ) WYETH.  
 XX  
 PI O'toolle M, Mounts WM, Shojaae N,  
 XX  
 DR WPI; 2004-441163/41.  
 XX  
 XX New isolated BFLP0169 nucleic acids and proteins, useful for diagnosing  
 PT or treating autoimmune disorder, e.g. lupus or lupus nephritis.  
 XX  
 PS Example 8; Page 67; 83pp; English.  
 XX  
 XX The present invention relates to human BFLP0169 (ADO71763) and its coding  
 CC sequence (ADO71762). BFLP0169 gene expression is increased in kidney  
 CC tissue in mice with lupus nephritis. The BFLP0169 sequences are useful  
 CC for diagnosing and treating autoimmune disorder, preferably lupus or  
 CC lupus nephritis. They are also useful in the generation of antibodies  
 CC that bind immunospecifically to the nucleic acids. The present sequence  
 CC is a BFLP0169 variant.  
 XX  
 SO Sequence 1753 AA;  
 Query Match 100.0%; Score 14; DB 8; Length 1753;  
 Best Local Similarity 100.0%; Pred. No. 3.9e-06;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MIRSKITSVLSFC 14  
 DB 1 MIRSKITSVLSFC 14  
 RESULT 16  
 ADO71774  
 ID ADO71774 standard; protein; 1753 AA.  
 XX  
 AC ADO71774;  
 XX  
 DT 26-AUG-2004 (first entry)  
 XX  
 DE BFLP0169 variant K1084R, SEQ ID 13.  
 XX  
 KM Antiinflammatory; Dermatological; Immunosuppressive; human; BFLP0169;  
 KM kidney; lupus nephritis; autoimmune disorder; lupus; mutant; mutain.  
 XX  
 OS Homo sapiens.  
 OS Synthetic.  
 XX  
 XX Key Location/Qualifiers  
 FT Misc-difference 1084 /note= "Wild-type K replaced with R"  
 PT  
 XX  
 PN WO2004048521-A2.  
 XX  
 PD 10-JUN-2004.  
 XX  
 PF 21-NOV-2003; 2003WO-US037339.

XX  
 PR 21-NOV-2002; 2002US-0428094P.  
 XX  
 PA (AMHP ) WYETH.  
 XX  
 PI O'toolle M, Mounts WM, Shojaae N,  
 XX  
 DR WPI; 2004-441163/41.  
 XX  
 XX New isolated BFLP0169 nucleic acids and proteins, useful for diagnosing  
 PT or treating autoimmune disorder, e.g. lupus or lupus nephritis.  
 XX  
 PS Example 12; Page 69-70; 83pp; English.  
 XX  
 XX The present invention relates to human BFLP0169 (ADO71763) and its coding  
 CC sequence (ADO71762). BFLP0169 gene expression is increased in kidney  
 CC tissue in mice with lupus nephritis. The BFLP0169 sequences are useful  
 CC for diagnosing and treating autoimmune disorder, preferably lupus or  
 CC lupus nephritis. They are also useful in the generation of antibodies  
 CC that bind immunospecifically to the nucleic acids. The present sequence  
 CC is a BFLP0169 variant.  
 XX  
 SO Sequence 1753 AA;  
 Query Match 100.0%; Score 14; DB 8; Length 1753;  
 Best Local Similarity 100.0%; Pred. No. 3.9e-06;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MIRSKITSVLSFC 14  
 DB 1 MIRSKITSVLSFC 14  
 RESULT 17  
 ADO71779  
 ID ADO71779 standard; protein; 1753 AA.  
 XX  
 AC ADO71779;  
 XX  
 DT 26-AUG-2004 (first entry)  
 XX  
 DE BFLP0169 variant D1542Q, SEQ ID 18.  
 XX  
 KM Antiinflammatory; Dermatological; Immunosuppressive; human; BFLP0169;  
 KM kidney; lupus nephritis; autoimmune disorder; lupus; mutant; mutain.  
 XX  
 OS Homo sapiens.  
 OS Synthetic.  
 XX  
 XX Key Location/Qualifiers  
 FT Misc-difference 1542 /note= "Wild-type D replaced with Q"  
 PT  
 XX  
 PN WO2004048521-A2.  
 XX  
 PD 10-JUN-2004.  
 XX  
 PF 21-NOV-2003; 2003WO-US037339.  
 XX  
 PR 21-NOV-2002; 2002US-0428094P.  
 XX  
 PA (AMHP ) WYETH.  
 XX  
 PI O'toolle M, Mounts WM, Shojaae N,  
 XX  
 DR WPI; 2004-441163/41.  
 XX  
 XX New isolated BFLP0169 nucleic acids and proteins, useful for diagnosing  
 PT or treating autoimmune disorder, e.g. lupus or lupus nephritis.  
 XX  
 PS Example 17; Page 72-73; 83pp; English.  
 XX  
 CC The present invention relates to human BFLP0169 (ADO71763) and its coding

CC sequence (ADO71762). BFLP0169 gene expression is increased in kidney  
 CC tissue in mice with lupus nephritis. The BFLP0169 sequences are useful  
 CC for diagnosing and treating autoimmune disorder, preferably lupus or  
 CC lupus nephritis. They are also useful in the generation of antibodies  
 CC that bind immunospecifically to the nucleic acids. The present sequence  
 CC is a BFLP0169 variant.

XX  
 SQ Sequence 1753 AA;

Query Match 100.0%; Score 14; DB 8; Length 1753;  
 Best Local Similarity 100.0%; Pred. No. 3.9e-06;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MIRSKITSVLSFC 14  
 |||||  
 DB 1 MIRSKITSVLSFC 14

RESULT 18

ADO71763  
 ID ADO71763 standard; protein; 1753 AA.

XX ADO71763;

XX 26-AUG-2004 (first entry)

XX Human BFLP0169, SEQ ID 2.

XX Antinflammatory; Dermatological; Immunosuppressive; human; BFLP0169;

KW kidney; lupus nephritis; autoimmune disorder; lupus.

XX Homo sapiens.

XX WO2004048521-A2.

XX 10-JUN-2004.

XX 21-NOV-2003; 2003WO-US037339.

XX 21-NOV-2002; 2002US-0428094P.

XX (AMHP ) WYETH.

XX O'Loole M, Mounts WM, Shojaaee N;

XX WPI; 2004-441163/41.

XX N-PSDB; ADO71762.

XX New isolated BFLP0169 nucleic acids and proteins, useful for diagnosing  
 PT or treating autoimmune disorder, e.g. lupus or lupus nephritis.

XX Claim 9; Page 7-8; 83pp; English.

CC The present invention relates to human BFLP0169 (ADO71763) and its coding  
 CC sequence (ADO71762). BFLP0169 gene expression is increased in kidney  
 CC tissue in mice with lupus nephritis. The BFLP0169 sequences are useful  
 CC for diagnosing and treating autoimmune disorder, preferably lupus or  
 CC lupus nephritis. They are also useful in the generation of antibodies  
 CC that bind immunospecifically to the nucleic acids.

XX  
 SQ Sequence 1753 AA;

Query Match 100.0%; Score 14; DB 8; Length 1753;  
 Best Local Similarity 100.0%; Pred. No. 3.9e-06;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MIRSKITSVLSFC 14  
 |||||  
 DB 1 MIRSKITSVLSFC 14

RESULT 19  
 ADO71776

ID ADO71776 standard; protein; 1753 AA.

XX ADO71776;

XX 26-AUG-2004 (first entry)

XX BFLP0169 variant 11247V, SEQ ID 15.

XX Antinflammatory; Dermatological; Immunosuppressive; human; BFLP0169;  
 KW kidney; lupus nephritis; autoimmune disorder; lupus; mutant; mutant.

XX Homo sapiens.

XX Synthetic.

XX Key Location/Qualifiers  
 FT Misc-difference 1247  
 FT /note= "Wild-type I replaced with V"

XX WO2004048521-A2.

XX 10-JUN-2004.

XX 21-NOV-2003; 2003WO-US037339.

XX 21-NOV-2002; 2002US-0428094P.

XX (AMHP ) WYETH.

XX O'Loole M, Mounts WM, Shojaaee N;

XX WPI; 2004-441163/41.

XX New isolated BFLP0169 nucleic acids and proteins, useful for diagnosing  
 PT or treating autoimmune disorder, e.g. lupus or lupus nephritis.

XX Example 14; Page 70-71; 83pp; English.

CC The present invention relates to human BFLP0169 (ADO71763) and its coding  
 CC sequence (ADO71762). BFLP0169 gene expression is increased in kidney  
 CC tissue in mice with lupus nephritis. The BFLP0169 sequences are useful  
 CC for diagnosing and treating autoimmune disorder, preferably lupus or  
 CC lupus nephritis. They are also useful in the generation of antibodies  
 CC that bind immunospecifically to the nucleic acids. The present sequence  
 CC is a BFLP0169 variant.

XX  
 SQ Sequence 1753 AA;

Query Match 100.0%; Score 14; DB 8; Length 1753;  
 Best Local Similarity 100.0%; Pred. No. 3.9e-06;  
 Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MIRSKITSVLSFC 14  
 |||||  
 DB 1 MIRSKITSVLSFC 14

RESULT 20  
 ADH85577

ID ADH85577 standard; protein; 162 AA.

XX ADH85577;

XX 22-APR-2004 (first entry)

XX Enterococcus faecalis polypeptide #57.

XX Enterococcus faecalis infection; transcription regulatory element;  
 KW antibacterial.

XX Enterococcus faecalis.

XX US6617156-B1.

XX

PD 09-SEP-2003.  
 XX 13-AUG-1998; 98US-00134000.  
 XX 15-AUG-1997; 97US-0055778P.  
 XX (DOHC/) DOHCETTE-STAMM L. A.  
 PA (BUSH/) BUSH D.  
 XX Doucette-Stamm LA, Bush D;  
 PI WPI: 2003-895394/82.  
 DR N-PSDB; ADH82172.  
 XX New nucleic acid comprising a sequence encoding an *Enterococcus faecalis*  
 PT polypeptide, useful for preparing a composition for diagnosing or  
 PR treating *E. faecalis* infection.  
 XX Disclosure; SEQ ID NO 3462; 193pp; English.  
 XX The invention relates to *Enterococcus faecalis* polynucleotides and  
 CC polypeptides. The invention also relates to a recombinant expression  
 CC vector comprising a polynucleotide operably linked to a transcription  
 CC regulatory element, a cell comprising a recombinant vector, a method for  
 CC producing an *E. faecalis* polypeptide, an isolated nucleic acid comprising  
 CC a sequence not given in the specification, a recombinant vector  
 CC comprising the nucleic acid and a cell comprising the recombinant vector.  
 CC The polynucleotides can be used to detect the presence of *E. faecalis* in  
 CC a sample. The sequences are useful for preparing a composition for  
 CC diagnosing or treating *Enterococcus faecalis* infection. This sequence  
 CC represents an *E. faecalis* polypeptide of the invention.  
 XX Sequence 162 AA;  
 SQ  
 QY Query Match 50.0%; Score 7; DB 7; Length 162;  
 DB Best Local Similarity 100.0%; Pred. No. 11;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 6 KITSVLS 12  
 DB 129 KITSVLS 135  
 RESULT 21  
 AAB79589  
 ID AAB79589 standard; protein; 285 AA.  
 XX  
 AC AAB79589;  
 DT 30-APR-2001 (first entry)  
 DE Corynebacterium glutamicum SMP protein sequence SEQ ID NO:694.  
 XX  
 KM Corynebacterium glutamicum; carbon metabolism and energy production;  
 KM SMP protein; sugar metabolism and oxidative phosphorylation protein;  
 KM fine chemical production; organic acid; pyrimidine base; nucleoside;  
 KM nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside;  
 KM nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;  
 KM carbohydrate; aromatic compound; vitamin; cofactor; polyketide; enzyme;  
 KM diagnosis; Corynebacterium diptheriae; evolutionary study.  
 XX  
 OS Corynebacterium glutamicum.  
 XX  
 PN WO200100844-A2.  
 XX  
 PD 04-JAN-2001.  
 PF 23-JUN-2000; 2000MO-IB000943.  
 XX  
 PR 25-JUN-1999; 99US-0141031P.  
 PR 08-JUL-1999; 99DE-01031412.  
 PR 08-JUL-1999; 99DE-01031413.  
 PR 08-JUL-1999; 99DE-01031419.  
 PR

PR 08-JUL-1999; 99DE-01031420.  
 PR 08-JUL-1999; 99DE-01031424.  
 PR 08-JUL-1999; 99DE-01031428.  
 PR 08-JUL-1999; 99DE-01031431.  
 PR 08-JUL-1999; 99DE-01031433.  
 PR 08-JUL-1999; 99DE-01031434.  
 PR 08-JUL-1999; 99DE-01031510.  
 PR 08-JUL-1999; 99DE-01031562.  
 PR 08-JUL-1999; 99DE-01031634.  
 PR 09-JUL-1999; 99DE-01032180.  
 PR 09-JUL-1999; 99DE-01032227.  
 PR 09-JUL-1999; 99DE-01032230.  
 PR 09-JUL-1999; 99US-0143208P.  
 PR 14-JUL-1999; 99DE-01032924.  
 PR 14-JUL-1999; 99DE-01032973.  
 PR 14-JUL-1999; 99DE-01033005.  
 PR 27-AUG-1999; 99DE-01040765.  
 PR 31-AUG-1999; 99US-0151572P.  
 PR 03-SEP-1999; 99DE-01042076.  
 PR 03-SEP-1999; 99DE-01042079.  
 PR 03-SEP-1999; 99DE-01042086.  
 PR 03-SEP-1999; 99DE-01042087.  
 PR 03-SEP-1999; 99DE-01042088.  
 PR 03-SEP-1999; 99DE-01042095.  
 PR 03-SEP-1999; 99DE-01042123.  
 PR 03-SEP-1999; 99DE-01042125.  
 XX (BADI ) BASF AG.  
 XX Pompejus M, Kroege B, Schroeder H, Zelder O, Habermayer G;  
 PI WPI: 2001-061975/07.  
 DR N-PSDB; AAF71706.  
 XX  
 DR New isolated Corynebacterium glutamicum nucleic acid encoding a sugar  
 PT metabolism and oxidative phosphorylation protein for production or  
 PT modulation of production of fine chemicals e.g. amino acids,  
 PT carbohydrates or enzymes.  
 XX  
 PS Claim 20; Page 1129-1130; 1246pp; English.  
 XX  
 CC AAF71360 to AAF71750 encode the Corynebacterium glutamicum sugar  
 CC metabolism and oxidative phosphorylation (SMP) proteins given in AAB79243  
 CC to AAB 79633 which are involved in carbon metabolism and energy  
 CC production. The C. glutamicum SMP gene can be used in vectors (II) for  
 CC expression in host cells and production or modulation of production of  
 CC fine chemicals, such as, an organic acid, a proteinogenic or  
 CC nonproteinogenic amino acid (preferred), a purine or pyrimidine base, a  
 CC nucleoside, a nucleotide, a lipid, a saturated or unsaturated fatty acid,  
 CC a diol, a carbohydrate, an aromatic compound, a vitamin, a cofactor, a  
 CC polyketide, or an enzyme. The presence of (I) or SMP proteins (III)  
 CC encoded by them are used for diagnosing the presence or activity of  
 CC Corynebacterium diptheriae in a subject. (I), (II), (III) or host cells  
 CC containing them are used to map genomes of organisms related to C.  
 CC glutamicum, identify and localise C. glutamicum sequences of interest, in  
 CC evolutionary studies, in determining SMP protein regions required for  
 CC function, in modulating SMP protein activity, in modulating the  
 CC metabolism of sugars, and in modulating high-energy molecule production  
 CC in a cell (I.e. ATP, NADPH)  
 CC  
 XX  
 SQ Sequence 285 AA;  
 QY Query Match 50.0%; Score 7; DB 4; Length 285;  
 DB Best Local Similarity 100.0%; Pred. No. 17;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 7 ITSVLSP 13  
 DB 139 ITSVLSP 145  
 RESULT 22  
 AAB67357

ID ABM67357 standard; protein; 366 AA.  
XX  
AC ABM67357;  
XX  
DT 20-NOV-2003 (first entry)  
XX  
DE Photorhabdus luminescens protein sequence #454.  
XX  
KM Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;  
KM detection; food; gene expression; plant; animal; microorganism; toxin;  
KM antibiotic; bioplastic; virulence factor; disease model; plague;  
KM whooping cough.  
XX  
OS Photorhabdus luminescens.  
XX  
FN WO200294867-A2.  
XX  
PD 28-NOV-2002.  
XX  
PF 07-FEB-2002; 2002WO-1B003040.  
XX  
PR 07-FEB-2001; 2001FR-00001659.  
XX  
PA (INSP) INST PASTEUR.  
XX (CNRS) CNRS CENT NAT RECH SCI.  
XX  
PI Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A,  
PI Buchrieser C;  
XX  
DR WPI; 2003-146459/14.  
XX  
PT Genomic sequence of Photorhabdus luminescens and encoded polypeptides,  
PT useful e.g. as therapeutic antimicrobials and agricultural pesticides.  
XX  
PS Claim 2; SEQ ID NO 454; 1205bp; French.  
XX  
CC The invention relates to the isolation of genes and their encoded  
CC proteins from Photorhabdus luminescens. The isolated sequences are  
CC sources of probes and primers for detecting the genome of P. luminescens  
CC and related species; to study polymorphisms; for gene analysis and for  
CC detection/amplification of the genes. Antibodies (Ab) raised against the  
CC polypeptides encoded by the genes are used for detection/identification  
CC of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that  
CC carry a gene-containing vector are used to select compounds that  
CC modulate, regulate, induce or inhibit expression of the genes in plants,  
CC animals or microorganisms other than P. luminescens and are able to alter  
CC response or sensitivity to toxins and antibiotics produced by P.  
CC luminescens. Cells transformed to express the genes are useful for  
CC recombinant production of the proteins, particularly toxins and  
CC antibacterials useful as insecticides, bactericides and fungicides. The  
CC genes, proteins, vectors containing the genes and Ab are also useful  
CC therapeutically (to treat microbial infection by bacteria or fungi that  
CC are sensitive to P. luminescens-encoded toxins or antibiotics) and as  
CC bioplastic. Other uses of the genes and the proteins are as virulence  
CC factors and for identifying targets of human diseases for which P.  
CC luminescens is a model (particularly plague and whooping cough). This  
CC sequence represents one of the isolated P. luminescens proteins  
XX  
SQ Sequence 366 AA:  
Query Match 50.0%; Score 7; DB 6; Length 366;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

AC ADO25476;  
XX  
DT 12-AUG-2004 (first entry)  
XX  
DE E.faecalis adenylsuccinate lyase purB protein Seg151.  
XX  
KM antimicrobial; microbial disease; drug composition; vaccine;  
KM bacterial infection; antibacterial; food preservative.  
XX  
OS Enterococcus faecalis.  
XX  
FN WO2004041854-A2.  
XX  
PD 21-MAY-2004.  
XX  
PF 05-NOV-2003; 2003WO-CA001671.  
XX  
PR 05-NOV-2002; 2002US-0423757P.  
PR 05-NOV-2002; 2002US-0423758P.  
PR 05-NOV-2002; 2002US-0423791P.  
PR 05-NOV-2002; 2002US-0423832P.  
PR 05-NOV-2002; 2002US-0423875P.  
PR 05-NOV-2002; 2002US-0423915P.  
PR 06-NOV-2002; 2002US-0424362P.  
PR 06-NOV-2002; 2002US-0424367P.  
PR 06-NOV-2002; 2002US-0424370P.  
PR 06-NOV-2002; 2002US-0424373P.  
PR 06-NOV-2002; 2002US-0424376P.  
PR 06-NOV-2002; 2002US-0424389P.  
PR 07-NOV-2002; 2002US-0424502P.  
PR 07-NOV-2002; 2002US-0424651P.  
PR 07-NOV-2002; 2002US-0424664P.  
PR 07-NOV-2002; 2002US-0424665P.  
PR 08-NOV-2002; 2002US-0424968P.  
PR 08-NOV-2002; 2002US-0425076P.  
PR 08-NOV-2002; 2002US-0425085P.  
PR 08-NOV-2002; 2002US-0425118P.  
PR 08-NOV-2002; 2002US-0425125P.  
PR 08-NOV-2002; 2002US-0425162P.  
PR 08-NOV-2002; 2002US-0425201P.  
PR 12-MAR-2003; 2003US-0453914P.  
PR 12-MAR-2003; 2003US-0454021P.  
PR 12-MAR-2003; 2003US-0454128P.  
PR 12-MAR-2003; 2003US-0454193P.  
PR 13-MAR-2003; 2003US-0454215P.  
PR 13-MAR-2003; 2003US-0454218P.  
PR 13-MAR-2003; 2003US-0454487P.  
PR 13-MAR-2003; 2003US-0454507P.  
PR 13-MAR-2003; 2003US-0454536P.  
PR 14-MAR-2003; 2003US-0455010P.  
PR 14-MAR-2003; 2003US-0455036P.  
PR 14-MAR-2003; 2003US-0455054P.  
PR 14-MAR-2003; 2003US-0455082P.  
PR 17-MAR-2003; 2003US-0455191P.  
PR 17-MAR-2003; 2003US-0455192P.  
PR 17-MAR-2003; 2003US-0455334P.  
PR 17-MAR-2003; 2003US-0455335P.  
PR 17-MAR-2003; 2003US-0455343P.  
XX  
XX (AFPI-) AFFINITUM PHARM INC.  
XX  
XX Edwards A, Dharmas A, Vedadi M, Domagala M, Nethery K;  
XX Mansoury K, Pinder B, Alam MZ, Ng I, Virag C, Houston S;  
XX McDonald M, Buzadzija K;  
XX  
XX WPI; 2004-400642/37.  
XX N-PSDB; ADO25475.  
XX  
XX Bacterial polypeptide composition useful for treating bacterial  
XX infection, has isolated, recombinant bacterial polypeptide such as GTP-  
XX binding protein Bra from Pseudomonas aeruginosa or adenylsuccinate lyase  
XX from Enterococcus faecalis.

PS Claim 40; SEQ ID NO 151; 566bp; English.

CC This invention relates to the identification of novel protein targets for  
 CC the development of antimicrobial drugs against pathogenic bacteria.  
 CC Specifically, it refers to recombinant proteins derived from  
 CC *Staphylococcus aureus*, *Helicobacter pylori*, *Streptococcus pneumoniae*,  
 CC *Bacteriella coli*, *Enterococcus faecalis* or *Pseudomonas aeruginosa*. The  
 CC present invention describes providing a three-dimensional structure for  
 CC these crystallised proteins to identify a potential modulator for the  
 CC prevention or treatment of microbial diseases. Furthermore, contacting a  
 CC protein with a modulator in drug composition or vaccine. Accordingly, such  
 CC compositions can be useful for treating bacterial infections, developing  
 CC antibacterial agents useful as food preservatives or treating food  
 CC products to eliminate potential pathogens. This polypeptide sequence is a  
 CC bacterial protein target of the invention.

SO Sequence 430 AA;

Query Match 50.0%; Score 7; DB 8; Length 430;  
 Best Local Similarity 100.0%; Pred. No. 25;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 KITSVLS 12  
 |||||  
 Db 397 KITSVLS 403

RESULT 24

AAU35142  
 ID AU35142 standard; protein; 431 AA.

XX AAU35142;

DT 13-FEB-2002 (first entry)

DE Enterococcus faecalis cellular proliferation protein #429.

XX Antisense; prokaryotic cellular proliferation protein; antibiotic;

KW antibacterial; drug design.

XX Enterococcus faecalis.

OS WO200170955-A2.

PN WO200170955-A2.

XX 27-SEP-2001.

PF 21-MAR-2001; 2001WO-US009180.

PR 21-MAR-2001; 2000US-0191078P.

PR 23-MAY-2000; 2000US-0206848P.

PR 26-MAY-2000; 2000US-0207727P.

PR 23-OCT-2000; 2000US-0242578P.

PR 27-NOV-2000; 2000US-0253625P.

PR 22-DEC-2000; 2000US-0257931P.

PR 16-FEB-2001; 2001US-0269308P.

XX (ELIT-) ELITRA PHARM INC.

PA (ELIT-) ELITRA PHARM INC.

PI Haasebeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ,

PI Yamamoto RT, Xu HH;

XX WPI; 2001-611495/70.

DR N-PSDB; AAS53001.

XX

PT New polynucleotides for the identification and development of

PT antibiotics, comprise sequences of antisense nucleic acids.

XX Example 3; SEQ ID NO 10735; 511bp; English.

CC The invention relates to antisense inhibitors of genes essential to

CC prokaryotic cellular proliferation, their use in identifying the genes,

CC their use in the discovery of novel antibiotics, the essential genes

CC themselves and the encoded proteins. The prokaryotes used are *Bacteriella*  
 CC *coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella pneumoniae*,  
 CC *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The invention is also  
 CC useful for the identification of potential new targets for antibiotic  
 CC development. The antisense nucleic acids can also be used to identify  
 CC proteins used in proliferation, to express these proteins, and to obtain  
 CC antibodies capable of binding to the expressed proteins. The proteins can  
 CC be used to screen compounds in rational drug discovery programmes. The  
 CC antisense nucleic acid sequence is also useful to screen for homologous  
 CC nucleic acids which are required for cell proliferation in a wide variety  
 CC of organisms. The present sequence represents an essential prokaryotic  
 CC cellular proliferation protein. Note: The sequence data for this patent  
 CC did not form part of the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp://wipo.int/pub/published\_pct\_sequences

SO Sequence 431 AA;

Query Match 50.0%; Score 7; DB 4; Length 431;  
 Best Local Similarity 100.0%; Pred. No. 25;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 KITSVLS 12  
 |||||  
 Db 398 KITSVLS 404

RESULT 25

ABU29114  
 ID ABU29114 standard; protein; 431 AA.

XX ABU29114;

DT 19-JUN-2003 (first entry)

DE Protein encoded by Prokaryotic essential gene #14641.

XX Antisense; prokaryotic essential gene; cell proliferation; drug design.

XX Enterococcus faecalis.

OS WO200277183-A2.

PN WO200277183-A2.

XX 03-OCT-2002.

PF 21-MAR-2002; 2002WO-US009107.

PR 21-MAR-2001; 2001US-00815242.

PR 06-SEP-2001; 2001US-00948993.

PR 25-OCT-2001; 2001US-0342923P.

PR 08-FEB-2002; 2002US-00072851.

PR 06-MAR-2002; 2002US-0362699P.

XX (ELIT-) ELITRA PHARM INC.

PA (ELIT-) ELITRA PHARM INC.

PI Wang L, Zamudio C, Malone C, Haasebeck R, Ohlsen KL, Zyskind JW,

PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Foreyth RA, Xu HH;

XX WPI; 2003-028926/02.

DR N-PSDB; AKA32984.

XX

PT New antisense nucleic acids, useful for identifying proteins or screening

PT for homologous nucleic acids required for cellular proliferation to

PT isolate candidate molecules for rational drug discovery programs.

PS Claim 25; SEQ ID NO 57038; 1766bp; English.

CC The invention relates to an isolated nucleic acid comprising any one of

CC the 6213 antisense sequences given in the specification where expression

CC of the nucleic acid inhibits proliferation of a cell. Also included are:

CC (1) a vector comprising a promoter operably linked to the nucleic acid

CC encoding a polypeptide whose expression is inhibited by the antisense

CC nucleic acid; (2) a host cell containing the vector; (3) an isolated

polypeptide or its fragment whose expression is inhibited by the antisense nucleic acid, (4) an antibody capable of specifically binding the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular proliferation or the activity of a gene in an operon required for proliferation; (7) identifying a compound that influences the activity of the gene product or that has an activity against a biological pathway required for proliferation, or that inhibits cellular proliferation; (8) identifying a gene required for cellular proliferation or the biological pathway in which a proliferation-required gene or its gene product lies or a gene on which the test compound that inhibits proliferation of an organism acts; (9) manufacturing an antibiotic; (10) profiling a compound's activity; (11) a culture comprising strains in which the product is overexpressed or underexpressed; (12) determining the extent to which each of the strains is present in a culture or collection of strains; or (13) identifying the target of a compound that inhibits the proliferation of an organism. The antisense nucleic acids are useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs, or for screening homologous nucleic acids required for proliferation in cells other than *S. aureus*, *S. typhimurium*, *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of the target prokaryotic essential genes. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences)

**Sequence 431 AA;**

Query Match	50.0%	Score 7;	DB 6;	Length 431;
Best Local Similarity	100.0%	Pred. No. 25;		
Matches 7;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

```
QY      6 KITSVLS 12
        |||||
Db     398 KITSVLS 404
```

RESULT 26  
ADO25474  
ID ADO25474 standard; protein; 431 AA

AC	ADO25474;
XX	
DT	12-AUG-2004 (first entry)

DE E<sub>-</sub>faecalis adenylosuccinate lyase purB protein Seg149.

KM antimicrobial; microbial disease; drug composition; vaccine; bacterial infection; antibacterial; food preservative.

OS Enterococcus faecalis.

PN WO2004041854-A2.

PD 21-MAY-2004

PF 05-NOV-2003; 2003WO-CA001671.

PR 05-NOV-2002; 2002US-0423757P. DE NOV 2003 0423757P

PR 05-NOV-2002; 2002US-0423791P.  
DE-NOV-2002; 2002US-0423832P

PR	05-NOV-2002: 2002US-0423915P
PR	03-NOV-2002: 2002US-0423875P

06-NOV-2002; 2002US-0424367P

PR 06-NOV-2002; 2002US-0424373P.

PR 06-NOV-2002; 2002US-0424389P.

PR 07-NOV-2002; 2002US-0424651P.

PR	07-NOV-2002	2002US-0424665P
PR	08-NOV-2002	2002US-0424968P
PR-	08-NOV-2002	2002US-0425076P
PR	08-NOV-2002	2002US-0425085P
PR	08-NOV-2002	2002US-0425118P
PR	08-NOV-2002	2002US-0425126P
PR	08-NOV-2002	2002US-0425162P
PR	08-NOV-2002	2002US-0425201P
PR	12-MAR-2003	2003US-0453914P
PR	12-MAR-2003	2003US-0454021P
PR	12-MAR-2003	2003US-0454128P
PR	12-MAR-2003	2003US-0454193P
PR	13-MAR-2003	2003US-0454215P
PR	13-MAR-2003	2003US-0454218P
PR	13-MAR-2003	2003US-0454387P
PR	13-MAR-2003	2003US-0454507P
PR	13-MAR-2003	2003US-0454536P
PR	14-MAR-2003	2003US-0455010P
PR	14-MAR-2003	2003US-0455036P
PR	14-MAR-2003	2003US-0455054P
PR	14-MAR-2003	2003US-0455082P
PR	17-MAR-2003	2003US-0455191P
PR	17-MAR-2003	2003US-0455192P
PR	17-MAR-2003	2003US-0455334P
PR	17-MAR-2003	2003US-0455335P
PR	17-MAR-2003	2003US-0455343P
XX		
PA	(AEPF-) AFFINIUM PHARM INC.	

PS Claim 40; SEQ ID NO 149; 566pp; English.

CC This invention relates to the identification of novel protein targets for  
CC the development of antimicrobial drugs against pathogenic bacteria.  
CC Specifically, it refers to recombinant proteins derived from  
CC *Staphylococcus aureus*, *Helicobacter pylori*, *Streptococcus pneumoniae*,  
CC *Escherichia coli*, *Enterococcus faecalis* or *Pseudomonas aeruginosa*. The  
CC present invention describes providing a three-dimensional structure for  
CC these crystallised proteins to identify a potential modulator for the  
CC prevention or treatment of microbial diseases. Furthermore, contacting a  
CC protein with a modulator can be useful for assaying protein activity and  
CC hence its viability in drug composition or vaccine. Accordingly, such  
CC compositions can be useful for treating bacterial infections, developing  
CC antibacterial agents useful as food preservatives or treating food  
CC products to eliminate potential pathogens. This polypeptide sequence is a  
CC bacterial protein target of the invention.

SQ Sequence 431 AA;

Query Match	50.0%	Score 7;	DB 8;	Length 431;
Best Local Similarity	100.0%	Pred. No. 25;		
Matches	7;	Conservative 0;	Mismatches 0;	Indels 0;
				Gaps 0

QY	6	KITSVLS	12
Db	398	KITSVLS	404

RESULT 27  
AAG92745  
ID AAG92745 standard; protein; 1019 AA

# XX





```

PN DE1054179-A1.
XX
XX 08-MAY-2003.
XX
XX 05-NOV-2001; 2001DE-01054179.
XX
XX 05-NOV-2001; 2001DE-01054179.
XX
XX (BADI ) BASF AG.
XX
XX Zelder O, Pompejus M, Schroeder H, Kroegeer B, Kloppeyge C,
PI Habehauer G;
XX
XX WPI; 2003-442553/42.
XX
XX N-PBDB; ADL65624.
XX
XX New nucleic acid encoding variant forms of membrane construction and
PT transport proteins, useful for production of fine chemicals from
PT microorganisms, e.g. nucleotides, lipids and especially lysine.
XX
XX Claim 1; Page; 20pp; German.
XX
XX This invention describes novel polynucleotides that encode membrane
CC synthase and membrane transport proteins from Corynebacterium glutamicum.
CC The polynucleotides are isolated from a nucleic acid library of C.
CC glutamicum then mutated at the specified positions, cloned and expressed
CC by standard methods. Cells, especially Corynebacterium glutamicum,
CC containing vectors that express the polynucleotides are used for
CC production of fine chemicals, preferably amino acids and specifically
CC lysine, but more generally nucleotides, nucleosides, lipids, fatty acids,
CC diols, carbohydrates, aromatic compounds, vitamins, co-factors and
CC enzymes. These are useful in the food, animal feed, cosmetics and
CC pharmaceutical industries. The polynucleotides, optionally as primers and
CC probes, can also be used for identification and classification of C.
CC glutamicum and related species, e.g. for diagnosis, for genomic mapping,
CC functional or evolutionary studies, gene manipulation and modulation of
CC metabolic activity. Cells that containing the polynucleotides of the
CC invention may produce fine chemicals in better yields, with higher
CC productivity and/or more efficiently. NOTE: This sequence is not
CC represented in the printed specification but is available in electronic
CC format. The sequence represented in this record has been obtained from
CC WO2003040292.
XX
XX
SQ Sequence 1019 AA;
Query Match 50.0%; Score 7; DB 7; Length 1019;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 7 ITSVLSF 13
|||||
Db 139 ITSVLSF 145
|||||
RESULT 30
AAB59995
ID AAB59995 standard; peptide; 12 AA.
XX
XX AAB59995;
XX
XX 05-NOV-2001 (first entry)
XX
XX Internalising peptide SEQ ID NO. 10.
XX
XX Internalising peptide; transport; apoptosis; arthritis; cancer;
XX stem cell; cell differentiation; immune response stimulation;
XX HIV vaccine.
XX
XX Synthetic.
XX
XX WO200115511-A2.
XX
XX 08-MAR-2001.

```

```

XX
XX 31-AUG-2000; 2000MO-US024034.
XX
XX 01-SEP-1999; 99US-0151980P.
XX
XX 13-MAR-2000; 2000US-0188944P.
XX
XX (UYP1-) UNIV PITTSBURGH.
XX
XX Robbins PD, M1 Z, Fritzell R, Glorioso JC, Gambotto A;
PI WPI; 2001-273309/28.
XX
XX Peptides that facilitate uptake and cytoplasmic and/or nuclear transport
PT of proteins, DNA and viruses, useful, e.g. for facilitating uptake of
PT antigens in immunogenic compositions.
XX
XX Claim 1; Page 115; 129pp; English.
XX
XX The present invention provides the sequences of 75 peptides which
CC facilitate the uptake and transport of viruses, proteins and nucleic
CC acids. These internalising peptides can be used for transport into the
CC cytoplasm or the nucleus. They are useful for facilitating uptake into
CC the cell, inducing apoptosis, for example in the treatment of arthritis
CC and cancer, to expand a population of stem cells or differentiated cells,
CC to stimulate cell differentiation, facilitate the integration of AAV into
CC the genome of a cell, and to stimulate an immune response, for example in
CC the case of a HIV vaccine. The present sequence is one of the peptides of
CC the invention.
XX
XX
SQ Sequence 12 AA;
Query Match 42.9%; Score 6; DB 4; Length 12;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 2 IRKSKI 7
|||||
Db 3 IRKSKI 8
|||||
RESULT 31
ADA86830
ID ADA86830 standard; peptide; 12 AA.
XX
XX ADA86830;
XX
XX 20-NOV-2003 (first entry)
XX
XX Internalised peptide pep10 SEQ ID NO:10.
XX
XX Internalising peptide; cytostatic; antiinflammatory; immunomodulator;
XX antiarthritic; cytoplasmic transport; nuclear transport;
XX peptide-cargo complex; apoptosis; arthritic; tumour; differentiation;
XX immune response; vaccine; inflammation; necrosis; transplantation;
XX cystic fibrosis; lung inflammation; gene therapy.
XX
XX Synthetic.
XX
XX WO2003068942-A2.
XX
XX 21-AUG-2003.
XX
XX 12-FEB-2003; 2003WO-US004632.
XX
XX 13-FEB-2002; 2002US-00075869.
XX
XX (UYP1-) UNIV PITTSBURGH.
XX
XX Robbins PD, M1 Z, Fritzell R, Glorioso JC, Gambotto A, Mai JC;
PI WPI; 2003-697526/66.
XX
XX New internalising peptides, useful for facilitating the delivery, uptake
PT

```

PT and cytoplasmic and/or nuclear transport of proteins, DNA or viruses into  
 PT a target cell, for inducing apoptosis in arthritic or tumor cells, or in  
 PT gene therapy.

PS Example 3, Page 18, 171pp, English.

CC The present invention describes an internalising peptide (I) comprising  
 CC any one of 14 fully defined amino acid sequences (designated PI-P14, see  
 CC ADA8896 to ADA8906, and ADA8917 to ADA8919). (I) has cytosolic,  
 CC anti-inflammatory, immunomodulator and antirheumatic activities.  
 CC Internalising peptides are useful for facilitating the delivery, uptake  
 CC and cytoplasmic and/or nuclear transport of cargo, e.g. proteins, DNA or  
 CC viruses, into a target cell. The internalising peptides and peptide-cargo  
 CC complexes in cells (e.g. arthritic cells or tumour cells), expanding a  
 CC population of stem cell or differentiated cells, stimulating the  
 CC differentiation of a population of stem cells, facilitating the  
 CC integration of adeno-associated virus DNA into the genome of a cell,  
 CC stimulating or eliciting an immune response in a subject, facilitating  
 CC the delivery of immunogens (e.g. vaccines), inhibiting the inflammatory  
 CC process, protecting tissue from apoptosis or necrosis during tissue  
 CC isolation prior to transplantation, facilitating transfer of proteins and  
 CC peptides to the lung for the treatment of cystic fibrosis or lung  
 CC inflammation, or in gene therapy. The present sequence represents a  
 CC peptide used in the exemplification of the present invention.

CC Sequence 12 AA:

Query Match 42.9%; Score 6; DB 7; Length 12;

Best Local Similarity 100.0%; Pred. No. 12;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 IRKSKI 7  
 |||||  
 Db 3 IRKSKI 8

RESULT 32

ABG78459  
 ID ABG78459 standard; protein; 17 AA.

AC ABG78459;

DT 15-NOV-2002 (first entry)

DE HLRSII, transmembrane domain #3.

XX Human; human leucine-rich repeat small intestine I; HLRSII; asthma;  
 KW proliferative disorder; gastrointestinal disorder; renal disorder;  
 KW neural disorder; reproductive disorder; calcium regulation; apoptosis;  
 KW immune system; anaemia; human immune deficiency virus; HIV; cancer;  
 KW blood coagulation disorder; autoimmune disorder; allergic reaction;  
 KW inflammatory condition; cardiovascular disorder; ischaemia;  
 KW neurological disorder; infectious disease; cytokine production;  
 KW expressed sequence tag; EST.

OS Homo sapiens.

PN WO200261086-A2.

PD 08-AUG-2002.

PF 20-DEC-2001; 2001WO-US049739.

PR 22-DEC-2000; 2000US-0257774P.

PA (BRIM ) BRISTOL-MYERS SQUIBB CO.

PI Feder J, Ramanathan C, Mintier G;

DR WPI, 2002-619252/66.

XX New isolated nucleic acid molecules encoding HLRSII polypeptides, or

PT their fragments and homologues, useful for preventing, treating and  
 PT ameliorating medical conditions, e.g. proliferative, gastrointestinal, or  
 PT renal disorders.

PS Disclosure; Page 225, 336pp, English.

CC The invention relates to isolated nucleic acid molecules (I) encoding  
 CC human leucine-rich repeat small intestine I (HLRSII) polypeptides. The  
 CC nucleic acid molecules and polypeptides are useful for preventing,  
 CC treating and ameliorating medical conditions, such as proliferative,  
 CC gastrointestinal, renal, neural, or reproductive disorders; or disorders  
 CC related to aberrant calcium regulation or apoptosis modulation, either  
 CC directly or indirectly. They are also useful for treating, preventing  
 CC and/or diagnosing diseases, disorders and/or conditions of: immune system  
 CC by activating or inhibiting the proliferation, differentiation, or  
 CC mobilisation of immune cells; haematopoietic cells e.g. thrombocytopenia,  
 CC anaemia; immunologic deficiency syndromes, e.g. human immune deficiency  
 CC virus (HIV) infection, HTLV-BLV infection; blood coagulation disorders,  
 CC e.g. arterial thrombosis; autoimmune disorders, e.g. Addison's disease,  
 CC myasthenia gravis; asthma or allergic reactions; inflammatory conditions,  
 CC e.g. chronic prostatitis, sepsis; proliferative disorders, e.g. cancer;  
 CC cardiovascular disorders, e.g. arrhythmia, myocardial ischaemia;  
 CC neuromuscular disorders, e.g. Alzheimer's disease, Huntington's  
 CC chorea; infectious diseases, e.g. measles, mumps, pneumonia, or viral,  
 CC bacterial, and fungal infections. The HLRSII polypeptides are useful for  
 CC modulating cytokine production, antigen presentation, or other processes  
 CC such as boosting immune responses. ABG78454-ABG78474 represent HLRSII  
 CC amino acid sequences and related amino acid sequences of the invention

CC Sequence 17 AA:

Query Match 42.9%; Score 6; DB 5; Length 17;

Best Local Similarity 100.0%; Pred. No. 17;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ITSVLS 12  
 |||||  
 Db 7 ITSVLS 12

RESULT 33

AAG00649  
 ID AAG00649 standard; protein; 77 AA.

AC AAG00649;

DT 06-OCT-2000 (first entry)

DE Human secreted protein, SEQ ID NO: 4730.

XX Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;  
 KW gene therapy; chromosome mapping.

OS Homo sapiens.

PN EP1033401-A2.

PD 06-SEP-2000.

PF 21-FEB-2000; 2000EP-00200610.

PR 26-FEB-1999; 99US-0122467P.

PA (GEST ) GENSET.

PI Dumas Milne Edwards J, Duclert A, Giordano J;

DR WPI, 2000-500381/45.

DR N-PSDB, AAC00655.

XX New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
 PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
 PT diagnostic, forensic, gene therapy and chromosome mapping procedures.

XX PS Claim 13; SEQ ID NO 4730; 71bp + Sequence Listing; English.

CC The present sequence is a polypeptide encoded by one of a large number of  
 CC 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs were  
 CC prepared from total human RNAs or polyA+ RNAs derived from 30 different  
 CC tissues. EST sequences usually correspond mainly to the 3' untranslated  
 CC region (UTR) of the mRNA because they are often obtained from oligo-dT  
 CC primed cDNA libraries. Such ESTs are not well suited for isolating cDNA  
 CC sequences derived from the 5' ends of mRNAs and even in those cases where  
 CC longer cDNA sequences have been obtained, the full 5' UTR is rarely  
 CC included. 5' ESTs are derived from mRNAs with intact 5' ends and can  
 CC therefore be used to obtain full length cDNAs and genomic DNAs. 5' ESTs  
 CC are also used in diagnostic, forensic, gene therapy and chromosome  
 CC mapping procedures. They are used to obtain upstream regulatory sequences  
 CC and to design expression and secretion vectors

CC XX

CC Sequence 77 AA;

Query Match 42.9%; Score 6; DB 3; Length 77;  
 Best Local Similarity 100.0%; Pred. No. 62;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 KSKITS 9  
 |||||  
 67 KSKITS 72

DB

RESULT 34  
 ABG06877  
 ID ABG06877 standard; protein; 85 AA.

XX AC ABG06877;  
 XX DT 13-FEB-2002 (first entry)  
 XX DE Novel human diagnostic protein #6868.  
 XX DE Human, chromosome mapping; gene mapping; gene therapy; forensic;  
 XX KW Food supplement; medical imaging; diagnostic; genetic disorder.  
 XX OS Homo sapiens.  
 XX PN WO200175067-A2.  
 XX PD 11-OCT-2001.  
 XX PF 30-MAR-2001; 2001WO-US008631.  
 XX PR 31-MAR-2000; 2000US-00540217.  
 XX PR 23-AUG-2000; 2000US-00649167.  
 XX PA (HYSE-) HYSEQ INC.  
 XX PI Dmanac RT, Liu C, Tang YF;  
 XX DR WPI; 2001-639362/73.  
 XX DR N-PSDB; AAS71064.

XX PT New isolated polynucleotide and encoded polypeptides, useful in  
 XX PT diagnostic, forensic, gene mapping, identification of mutations  
 XX PT responsible for genetic disorders or other traits and to assess  
 XX PT biodiversity.

XX PS Claim 20; SEQ ID NO 37236; 103bp; English.

CC The invention relates to isolated polynucleotide (I) and polypeptide (II)  
 CC sequences. (I) is useful as hybridisation probes, polymerase chain  
 CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
 CC in recombinant production of (II). The polynucleotides are also used  
 CC in diagnostics as expressed sequence tags for identifying expressed  
 CC genes. (II) is useful in gene therapy techniques to restore normal  
 CC activity of (II) or to treat disease states involving (II). (II) is

CC useful for generating antibodies against it, detecting or quantitating a  
 CC polypeptide in tissue, as molecular weight markers and as a food  
 CC supplement. (II) and its binding partners are useful in medical imaging  
 CC of sites expressing (II). (I) and (II) are useful for treating disorders  
 CC involving aberrant protein expression or biological activity. The  
 CC polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG037 represent novel human diagnostic  
 CC patent did not appear in the printed specification, but was obtained in  
 CC electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences

CC XX

CC Sequence 85 AA;

Query Match 42.9%; Score 6; DB 4; Length 85;  
 Best Local Similarity 100.0%; Pred. No. 67;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MIRKSK 6  
 |||||  
 49 MIRKSK 54

DB

RESULT 35  
 AAU39937  
 ID AAU39937 standard; protein; 90 AA.

XX AC AAU39937;  
 XX DT 13-FEB-2002 (first entry)  
 XX DE Propionibacterium acnes immunogenic protein #833.  
 XX DE Propionibacterium acnes immunogenic protein #833.  
 XX KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;  
 XX KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;  
 XX KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;  
 XX KW dermatological; osteopathic; neuroprotectant.  
 XX OS Propionibacterium acnes.  
 XX PN WO200181581-A2.  
 XX PD 01-NOV-2001.  
 XX PF 20-APR-2001; 2001WO-US012865.  
 XX PR 21-APR-2000; 2000US-0199047P.  
 XX PR 02-JUN-2000; 2000US-0208841P.  
 XX PR 07-JUL-2000; 2000US-0216747P.  
 XX PA (COR-) CORIXA CORP.  
 XX PI Skelky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;  
 XX PI L'maisonneuve J, Zhang Y, Jen S, Carter D;  
 XX DR WPI; 2001-616774/71.  
 XX DR N-PSDB; AAS59510.

XX PT Propionibacterium acnes polypeptides and nucleic acids useful for  
 XX PT vaccinating against and diagnosing infections, especially useful for  
 XX PT treating acne vulgaris.

XX PS Example 1; SEQ ID NO 1132; 1069bp; English.

CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic  
 CC polypeptides. The proteins and their associated DNA sequences are used in  
 CC the treatment, prevention and diagnosis of medical conditions caused by  
 CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,  
 CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.  
 CC P. acnes is also involved in infections of bone, joints and the central

CC nervous system, however it is particularly involved in the inflammatory  
CC lesions associated with acne vulgaris. A method for detecting the  
CC presence or absence of P. acnes in a patient comprises contacting the  
CC sample with a binding agent that binds to the proteins of the invention  
CC and determining the amount of bound protein in the sample. The  
CC polypeptides may be used as antigens in the production of antibodies  
CC specific for P. acnes proteins. These antibodies can be used to  
CC downregulate expression and activity of P. acnes polypeptides and  
CC therefore treat P. acnes infections. The antibodies may also be used as  
CC diagnostic agents for determining P. acnes presence, for example, by  
CC enzyme linked immunosorbent assay (ELISA). Note: The sequence data for  
CC this patent did not form part of the printed specification, but was  
CC obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

CC Sequence 90 AA;

Query Match 42.9%; Score 6; DB 4; Length 90;  
Best Local Similarity 100.0%; Pred. No. 70;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ITSVLS 12  
DB 5 ITSVLS 10

RESULT 36  
ABM36456  
ID ABM36456 standard; protein; 90 AA.  
XX  
AC ABM36456;  
XX  
DT 20-OCT-2003 (first entry)  
XX  
DE Propionibacterium acnes predicted ORF-encoded polypeptide #1132.  
XX  
KM Acne vulgaris; anti-seborrheic; dermatological; antibacterial;  
XX Immunostimulant; immune response; vaccine.  
XX  
OS Propionibacterium acnes.  
XX  
PN WO2003033515-A1.  
XX  
PD 24-APR-2003.  
XX  
PF 11-OCT-2002; 2002WO-US032727.  
XX  
PR 15-OCT-2001; 2001US-00978825.  
XX  
PA (CORI-) CORIXA CORP.  
XX  
PI Mitcham JL, Skeiky YAM, Persing DH, Bhatia A, Maisonneuve JL,  
XX Zhang Y, Wang S, Jen S, Lodes MT, Benson DR, Jones R, Carter D;  
XX Barth B, Valliave-Douglas J;  
XX  
DR WPI; 2003-381789/36.  
XX N-PSDB; ACF64439.  
XX

PT New Propionibacterium acnes polypeptides and polynucleotides encoding the  
PT polypeptide, useful for diagnosing, preventing or treating acne vulgaris,  
PT or for stimulating an immune response specific for a P. acnes protein.  
XX  
PS Example 1; SEQ ID NO 1132; 1481bp; English.

CC The invention relates to an isolated polynucleotide (ACF64435-ACF64733)  
CC encoding a Propionibacterium acnes protein. The invention also relates to  
CC polypeptides encoded by the polynucleotides (ABM35624-ABM64536) and to  
CC immunogenic fragments of P. acnes polypeptides. The invention  
CC additionally encompasses expression vectors and host cells comprising a  
CC polynucleotide of the invention; antibodies against polypeptides of the  
CC invention; fusion proteins comprising a polypeptide of the invention; a  
CC method for stimulating an immune response specific for a P. acnes  
CC polypeptide and an isolated T cell population comprising T cells prepared

CC via this method, a vaccine composition (comprising P. acnes polypeptides,  
CC polynucleotides, antibodies, fusion proteins, T cell populations, or  
CC antigen-presenting cells that express the polypeptide); a method and kit  
CC for detecting or determining the presence or absence of P. acnes in a  
CC patient; and a method for inhibiting the development of P. acnes in a  
CC patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion  
CC polypeptides are useful for diagnosing, preventing or treating acne  
CC vulgaris, or for stimulating an immune response specific for a P. acnes  
CC protein. The polynucleotides can also be used as probes or primers for  
CC nucleic acid hybridisation. The vaccine composition is useful for the  
CC stimulation of an immune response against P. acnes, or for treating acne,  
CC and the kit is useful for performing a diagnostic assay. The present  
CC sequence represents a polypeptide predicted to be encoded by an ORF (open  
CC reading frame) contained within the P. acnes polynucleotides of the  
CC invention. Note: The sequence data for this patent did not form part of  
CC the printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

CC Sequence 90 AA;

Query Match 42.9%; Score 6; DB 6; Length 90;  
Best Local Similarity 100.0%; Pred. No. 70;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ITSVLS 12  
DB 5 ITSVLS 10

RESULT 37  
ABU17550  
ID ABU17550 standard; protein; 130 AA.  
XX  
AC ABU17550;  
XX  
DT 19-JUN-2003 (first entry)  
XX  
DE Protein encoded by Prokaryotic essential gene #3077.  
XX  
KM Antisense; prokaryotic essential gene; cell proliferation; drug design.  
XX  
OS Bacillus anthracis.  
XX  
PN WO200277183-A2.  
XX  
PD 03-OCT-2002.  
XX  
PF 21-MAR-2002; 2002WO-US009107.  
XX  
PR 21-MAR-2001; 2001US-00815242.  
XX  
PR 06-SEP-2001; 2001US-00948993.  
XX  
PR 25-OCT-2001; 2001US-0349232P.  
XX  
PR 08-FEB-2002; 2002US-00072851.  
XX  
PR 06-MAR-2002; 2002US-0362699P.  
XX  
PA (ELIT-) ELITRA PHARM INC.  
XX  
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zykkind JW,  
XX Wall D, Traxwick JD, Carr GJ, Yamamoto R, Foreyth RA, Xu HH;  
XX  
DR WPI; 2003-029926/02.  
XX N-PSDB; ACA21420.  
XX

PT New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.  
XX  
PS Claim 25; SEQ ID NO 45474; 1766bp; English.

CC The invention relates to an isolated nucleic acid comprising any one of  
CC the 623 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:

(1) a vector comprising a promoter operably linked to the nucleic acid encoding a polypeptide whose expression is inhibited by the antisense nucleic acid; (2) a host cell containing the vector; (3) an isolated polypeptide or its fragment whose expression is inhibited by the antisense nucleic acid; (4) an antibody capable of specifically binding the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular proliferation or the activity of a gene in an operon required for proliferation; (7) identifying a compound that influences the activity of the gene product or that has an activity against a biological pathway required for proliferation, or that inhibits cellular proliferation; (8) identifying a gene required for cellular proliferation or the biological pathway in which a proliferation-required gene or its gene product lies or a gene on which the target compound that inhibits proliferation of an organism acts; (9) manufacturing an antibiotic; (10) profiling a compound's activity; (11) a culture comprising strains in which the gene product is overexpressed or underexpressed; (12) determining the extent to which each of the strains is present in a culture or collection of strains; or (13) identifying the target of a compound that inhibits the proliferation of an organism. The antisense nucleic acids are useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs, or for screening homologous nucleic acids required for proliferation in cells other than *S. aureus*, *S. typhimurium*, *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of the target prokaryotic essential genes. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic form directly from WIPO at [http://wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences)

Sequence 130 AA;

Query Match 42.9%; Score 6; DB 6; Length 130;  
Best Local Similarity 100.0%; Pred. No. 97;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 3 RKSKIT 8  
| | | | |  
Db 28 RKSKIT 33

RESULT 38  
ADCF94908  
ID ADC94908 standard; protein; 133 AA.  
XX  
AC ADC94908;  
XX  
DT 01-JAN-2004 (first entry)  
XX  
DE E. faecium protein sequence SEQ ID 4535.  
XX  
KW Vaccine; urinary tract infection; bacteraemia; endocarditis; wound; abdominal-pelvic infection.  
XX  
OS Enterococcus faecium.  
XX  
PN US6583275-B1.  
XX  
PD 24-JUN-2003.  
XX  
PF 30-JUN-1998; 98US-00107532.  
XX  
PR 02-JUL-1997; 97US-0051571P.  
PR 14-MAY-1998; 98US-0085598P.  
XX  
PA (GENO-) GENOME THERAPEUTICS CORP.  
XX  
PI Doucette-Stamm LA, Bush D;  
XX  
DR WPI, 2003-799836/75.  
DR N-PSDB; ADC91254.  
XX  
PT New isolated nucleic acid derived from Enterococcus faecium encoding an Enterococcus faecium polypeptide useful for detection, prevention and

PT treatment of a pathological condition resulting from a bacterial infection.

PS Example 1; SEQ ID NO 4535; 243pp; English.

XX The invention relates to an isolated nucleic acid derived from  
CC Enterococcus faecium encoding an Enterococcus faecium polypeptide having  
CC one of 10 fully defined sequences given in the (or comprising 40  
CC sequential nucleotides chosen from any of the nucleic acids, its  
CC complement or sequences hybridizing to it). Also included are a  
CC recombinant vector comprising the nucleic acid operably linked to  
CC transcription regulatory element, a cell comprising the vector and a  
CC single-stranded probe comprising the nucleic acid. The nucleic acids are  
CC chosen from 3654 disclosed sequences encoding 3654 disclosed proteins.  
CC The nucleic acids are useful for diagnosing pathological conditions  
CC resulting from E. faecium bacterial infection (e.g. urinary tract  
CC infection), bacteraemia, endocarditis, wounds and abdominal-pelvic  
CC infection) and for screening drugs such as agonists and antagonists. The  
CC nucleic acid is useful for recombinant production of *Candida albicans* -  
CC derived peptides or antisense polypeptides. Pharmaceutical compositions  
CC and vaccines containing the nucleic acid are useful for preventing or  
CC treating Enterococcus faecium infections. The present sequence represents  
CC one if the disclosed E. faecium proteins.

Sequence 133 AA;

Query Match 42.9%; Score 6; DB 7; Length 133;  
Best Local Similarity 100.0%; Pred. No. 99;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 KSKITS 9  
| | | | |  
Db 120 KSKITS 125

RESULT 39  
ADPF04124  
ID ADPF04124 standard; protein; 135 AA.  
XX  
AC ADPF04124;  
XX  
DT 12-FEB-2004 (first entry)  
XX  
DE Bacterial polypeptide #237.  
XX  
KW Proteus mirabilis infection; bacterial infection; antibacterial; immunostimulant.  
XX  
OS Proteus mirabilis.  
XX  
PN US6605709-B1.  
XX  
PD 12-AUG-2003.  
XX  
PF 05-APR-2000; 2000US-00543681.  
XX  
PR 09-APR-1999; 99US-0128706P.  
XX  
PA (GENO-) GENOME THERAPEUTICS CORP.  
XX  
PI Breton GL;  
XX  
DR WPI, 2003-895291/82.  
DR N-PSDB; ADE99951.  
XX  
PT New Proteus mirabilis polypeptides and polymucleotides, useful as reagents for diagnosis of bacterial disease, as components of antibacterial vaccines, as targets for antibacterial drugs, or as biocontrol agents for plants.  
XX  
PS Disclosure; SEQ ID NO 4409; 870pp; English.  
XX  
PT The invention relates to new Proteus mirabilis polypeptides and

CC polynucleotides. The invention also relates to antibodies against the  
 CC polypeptides, methods for producing the polypeptides, a method of  
 CC generating vaccines for immunising an individual against *P. mirabilis*, a  
 CC method for evaluating a compound for the ability to bind a *P. mirabilis*  
 CC polypeptide and a method for screening test compounds for anti-bacterial  
 CC activity. The polypeptides and polynucleotides are useful as molecular  
 CC targets for diagnosing, preventing and treating pathological conditions  
 CC resulting from bacterial infection, as reagents for diagnosis of  
 CC bacterial diseases, as components of antibacterial vaccines, as targets  
 CC for antibacterial drugs or as bio-control agents for plants. This  
 CC sequence represents a *Proteus mirabilis* polypeptide of the invention.

SO Sequence 135 AA;

Query Match 42.9%; Score 6; DB 7; Length 135;  
 Best Local Similarity 100.0%; Pred. No. 1e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 6 ITSVL 11  
 |||||  
 Db 3 ITSVL 8

RESULT 40  
 ABP10237  
 ID ABP10237 standard; protein; 139 AA.  
 XX  
 AC ABP10237;  
 XX  
 DT 24-JUN-2002 (first entry)  
 XX  
 DE Human ORFX protein sequence SEQ ID NO:20456.  
 XX  
 KM Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;  
 KM hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;  
 KM degenerative disorder; osteoarthritis; neurodegenerative disorder;  
 KM cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;  
 KM hypertension; hypothyroidism; cholesterol ester storage disease;  
 KM immune deficiency; immune disorder; infectious disease;  
 KM autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;  
 KM myasthenia gravis.  
 XX  
 OS Homo sapiens.  
 XX  
 PN MO200192523-A2.  
 XX  
 PD 06-DEC-2001.  
 XX  
 PE 29-MAY-2001; 2001WO-US010836.  
 XX  
 PR 30-MAY-2000; 2000US-0206132P.  
 PR 29-AUG-2000; 2000US-0228716P.  
 XX  
 PA (CURA-) CURAGEN CORP.  
 XX  
 PI Shimkets RA, Leach MD;  
 XX  
 DR WPI; 2002-106308/14.  
 DR N-PSDB; ABN25989.  
 XX  
 PT Novel human polypeptides and polynucleotides useful for diagnosing,  
 PT preventing and treating cardiovascular disease, neurodegenerative,  
 PT hyperproliferative disorders and autoimmune disorders.  
 XX  
 PS Disclosure; SEQ ID NO 20456; 1037bp; English.  
 XX  
 CC The present invention describes substantially purified human proteins  
 CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1  
 CC in the specification)). ABN15762 to ABN27252 encode the human ORFX  
 CC proteins given in ABP00010 to ABP1500. ORFX proteins are useful for  
 CC treating or preventing a pathology associated with an ORFX-associated  
 CC disorder in humans, and in the manufacture of a medicament for treating a  
 CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide

CC sequences can be used in gene therapy. ORFX sequences can be used in the  
 CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,  
 CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,  
 CC osteoarthritis, neurodegenerative disorders, disorders related to organ  
 CC transplantation, cardiovascular diseases, diabetes mellitus, systemic  
 CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester  
 CC storage disease, various immune deficiencies and disorders, infectious  
 CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid  
 CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host  
 CC disease and autoimmune inflammatory eye disease. ORFX proteins are also  
 CC useful for treating burns, incisions, ulcers, for treating osteoporosis,  
 CC bone degenerative disorders, or periodontal disease, and for gut  
 CC protection or regeneration and treatment of lung or liver fibrosis,  
 CC reperfusion injury in various tissues and conditions resulting from  
 CC systemic cytokine damage. N.B. The sequence data for this patent did not  
 CC form part of the printed specification, but was obtained in electronic  
 CC format directly from WIPO at ftp.wipo.int/pub/published\_pat\_sequences

SO Sequence 139 AA;

Query Match 42.9%; Score 6; DB 5; Length 139;  
 Best Local Similarity 100.0%; Pred. No. 1e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 7 ITSVL 12  
 |||||  
 Db 60 ITSVL 65

Search completed: December 30, 2004, 15:33:17  
 Job time : 165 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OW protein - protein search, using sw model

Run on: December 30, 2004, 14:45:25 / Search time 273.065 Seconds  
(without alignments)  
3693.737 Million cell updates/sec

Title: US-10-719-385-2

Perfect score: 9007

Sequence: 1 MRKSKITSTVLSFCRSRREL.....PESQEPILQVQAFVHMQR 1753

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 150 summaries

Database : Uniprot\_02:\*  
1: uniprot\_sprot:\*  
2: uniprot\_trembl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8939	99.2	1745	2 Q14675	014675 homo sapien
2	8482	94.2	1761	2 Q6Z0H8	06Z0H8 mus musculu
3	8482	94.2	1761	2 BAC97883	Bac97883 mus musculu
4	8270.5	91.8	1639	2 Q723K8	Q723K8 homo sapien
5	6608	73.4	1233	2 Q81W11	Q81W11 homo sapien
6	4093	45.4	853	2 Q9BS12	Q9BS12 mus musculu
7	3939	43.7	798	2 Q8C7A1	Q8C7A1 mus musculu
8	2420	26.9	525	2 Q8R3F1	Q8R3F1 mus musculu
9	1852	20.6	403	2 Q80U14	Q80U14 mus musculu
10	848	9.4	1056	2 Q7Q9K3	Q7Q9K3 anopheles g
11	829	9.2	190	2 Q89083	Q89083 mus musculu
12	801.5	8.9	1832	2 Q9V6F6	Q9V6F6 drosophila
13	372.5	4.1	1997	2 Q6DRN9	Q6DRN9 brachydanio
14	368.5	4.1	2025	2 Q8LW11	Q8LW11 oryza sativ
15	292.5	3.2	2012	1 N205_HUMAN	Q92621 homo sapien
16	292	3.2	1139	2 Q8S666	Q8S666 oryza sativ
17	291.5	3.2	2003	2 Q6PDG0	Q6PDG0 mus musculu
18	291.5	3.2	2003	2 AAH58729	AAH58729 mus musculu
19	289.5	3.2	2067	2 Q6ZOG1	Q6ZOG1 mus musculu
20	289.5	3.2	2067	2 BAC97903	Bac97903 mus musculu
21	286	3.2	1851	2 Q6P9U5	Q6P9U5 mus musculu
22	286	3.2	1851	2 AAH60139	AAH60139 mus musculu
23	286	3.0	1851	2 Q6BKR9	Q6BKR9 debaryomyce
24	269	3.0	1641	2 Q6BKR9	Q6BKR9 debaryomyce
25	258.5	2.9	1501	2 Q756X1	Q756X1 ashbya gos
26	258.5	2.9	1501	2 AA552965	AA552965 ashbya gos
27	258.5	2.9	1670	2 Q6P9U5	Q6P9U5 ashbya gos
28	233.5	2.6	1564	1 N184_SCHPO	Q9P7M8 schizosacch
29	232	2.6	1710	2 Q6P9U9	Q6P9U9 candida gla
30	218	2.4	2159	2 Q78Y48	Q78Y48 brachydanio
31	217.5	2.4	1818	2 Q6MYH8	Q6MYH8 aspergillus

32	217.5	2.4	1818	2 CAF32025	CAF32025 aspergill
33	215.5	2.4	561	2 Q9T014	Q9T014 arabidopsis
34	214	2.4	1655	1 N188_YEAST	N188_YEAST
35	212	2.4	1548	2 Q6C1T5	Q6C1T5
36	212	2.4	1606	2 Q6CNP1	Q6CNP1 kluyveromyc
37	205	2.3	1837	2 Q9LUS3	Q9LUS3 arabidopsis
38	197.5	2.2	1278	2 Q7T153	Q7T153 brachydanio
39	197.5	2.2	1336	2 Q7T152	Q7T152 brachydanio
40	197.5	2.2	2057	2 Q9VWB8	Q9VWB8 drosophila
41	197.5	2.2	2030	2 Q810V9	Q810V9 drosophila
42	192.5	2.1	1647	2 P78847	P78847 schizosacch
43	192.5	2.1	1683	1 N192_YEAST	N192_YEAST
44	189	2.1	1248	2 Q9VKY2	Q9VKY2 drosophila
45	184.5	2.0	2670	1 YAO5_SCHPO	YAO5_SCHPO
46	179.5	2.0	5280	2 Q7R0B5	Q7R0B5 giardia lam
47	175.5	1.9	1073	2 Q9T013	Q9T013 arabidopsis
48	175	1.9	2110	2 Q9CAQ9	Q9CAQ9 arabidopsis
49	171.5	1.9	2104	1 MY93_SCHPO	MY93_SCHPO
50	169.5	1.9	2443	2 Q96J17	Q96J17 homo sapien
51	169	1.9	1585	2 Q72745	Q72745 homo sapien
52	169	1.9	2658	2 Q9XIR5	Q9XIR5 homo sapien
53	167	1.9	4385	2 Q81G62	Q81G62 arabidopsi
54	166	1.8	2666	2 Q6FW99	Q6FW99 candida gla
55	163.5	1.8	2076	2 Q01485	Q01485 caenorhabd
56	162.5	1.8	2392	2 Q95001	Q95001 homo sapien
57	162	1.8	2444	2 Q6J3M3	Q6J3M3 aedes aegy
58	162	1.8	2615	2 Q7XPP8	Q7XPP8 oryza sativ
59	161	1.8	1858	2 Q7RXM6	Q7RXM6 neurospora
60	160	1.8	1830	1 BP28_ARATH	BP28_ARATH
61	160	1.8	1835	2 Q6BMT4	Q6BMT4 arabidopsi
62	159.5	1.8	2598	2 Q7RNO6	Q7RNO6 debaryomyce
63	158.5	1.8	2362	2 Q7T160	Q7T160 brachydanio
64	158.5	1.8	3616	2 Q9K6V0	Q9K6V0 gallus gall
65	158	1.8	2036	2 Q7Q0A2	Q7Q0A2 giardia lam
66	158	1.8	4283	2 Q8UW17	Q8UW17 fugu rubrip
67	157.5	1.7	1648	2 Q6CNK8	Q6CNK8 kluyveromyc
68	157.5	1.7	2607	2 Q6L4J5	Q6L4J5 solanum dem
69	157.5	1.7	2607	2 AA70487	AA70487 solanum d
70	157.5	1.7	2736	2 Q6BMO6	Q6BMO6 debaryomyce
71	157	1.7	2456	2 Q6FUD9	Q6FUD9 candida gla
72	157	1.7	3859	1 TRAP_HUMAN	TRAP_HUMAN
73	157	1.7	5058	2 Q86UQ4	Q86UQ4 homo sapien
74	156	1.7	835	2 Q81D88	Q81D88 drosophila
75	156	1.7	2605	2 Q9W246	Q9W246 drosophila
76	156	1.7	2733	2 Q81D88	Q81D88 drosophila
77	155	1.7	2432	1 TALA_DICDI	TALA_DICDI
78	154.5	1.7	2113	2 Q7Q3X7	Q7Q3X7 anopheles g
79	154.5	1.7	3227	1 PTR1_SCHPO	PTR1_SCHPO
80	154	1.7	2538	2 Q6BXX3	Q6BXX3 debaryomyce
81	154	1.7	2576	2 Q71A17	Q71A17 drosophila
82	154	1.7	2576	2 AAH60444	AAH60444 drosophila
83	154	1.7	3966	2 Q7KTP2	Q7KTP2 drosophila
84	154	1.7	3996	2 AAH10532	AAH10532 drosophila
85	153.5	1.7	2836	2 Q9U3B3	Q9U3B3 caenorhabd
86	153.5	1.7	2862	2 Q9Y0A2	Q9Y0A2 caenorhabd
87	153.5	1.7	2882	2 Q9Y0A1	Q9Y0A1 caenorhabd
88	153	1.7	4568	1 DYHC_CAEEL	DYHC_CAEEL
89	153	1.7	5635	2 Q9Y9N1	Q9Y9N1 leishmania
90	152	1.7	1196	2 Q9ATZ4	Q9ATZ4 linum usita
91	151.5	1.7	1445	2 Q6NMS6	Q6NMS6 brachydanio
92	151.5	1.7	1445	2 AAQ73628	AAQ73628 brachydanio
93	151.5	1.7	1920	2 Q6BNV7	Q6BNV7 debaryomyce
94	151	1.7	3259	1 G0B1_HUMAN	G0B1_HUMAN
95	150.5	1.7	1778	2 Q86XN9	Q86XN9 homo sapien
96	150.5	1.7	8407	2 Q7RTM4	Q7RTM4 homo sapien
97	150	1.7	1403	2 Q80U74	Q80U74 mus musculu
98	150	1.7	2847	2 Q95X74	Q95X74 caenorhabd
99	150	1.7	2903	2 Q95X75	Q95X75 caenorhabd
100	149.5	1.7	1334	2 Q9X7S7	Q9X7S7 caenorhabd
101	149.5	1.7	1998	2 Q8JFV4	Q8JFV4 giardia lam
102	149	1.7	946	2 Q7QYZ0	Q7QYZ0 brachydanio
103	149	1.7	1562	2 Q7QUB1	Q7QUB1 giardia lam
104	148.5	1.6	2675	2 Q92616	Q92616 homo sapien

105 147.5 1.6 1088 1 RP17 HUMAN  
106 147.5 1.6 1088 2 CAC1055  
107 147.5 1.6 2209 1 KXTC HUMAN  
108 147.5 1.6 2241 2 O8AM10  
109 147 1.6 2144 1 BP28 HUMAN  
110 147 1.6 2470 2 O9VK45  
111 147 1.6 2671 2 O754A3  
112 147 1.6 2671 2 AAS53540  
113 147 1.6 2954 2 O42263  
114 147 1.6 6885 1 SNEZ\_HUMAN  
115 146.5 1.6 1170 2 O8MSV6  
116 146.5 1.6 1868 1 YHDO\_YEAST  
117 146.5 1.6 2471 2 O6FNE7  
118 146.5 1.6 2509 2 O7S6V7  
119 146.5 1.6 2717 2 O7QJ3  
120 146.5 1.6 4829 1 BIR6\_HUMAN  
121 146 1.6 1441 2 O80TF2  
122 146 1.6 1200 2 Y539\_HUMAN  
123 145.5 1.6 1574 2 O8ATD1  
124 145.5 1.6 1629 2 O9BIR5  
125 145.5 1.6 2857 2 O9Y0A0  
126 145.5 1.6 3424 2 O9VT22  
127 145.5 1.6 3479 2 O9VT28  
128 145.5 1.6 3479 2 O9GPT8  
129 145.5 1.6 3479 2 O9GPT8  
130 145 1.6 1005 1 EVC\_MOUSE  
131 144.5 1.6 1545 2 O8BNO  
132 144.5 1.6 1728 2 O7R067  
133 144.5 1.6 2594 1 7LES\_DROVI  
134 144.5 1.6 5412 2 O7R3N4  
135 144 1.6 1339 2 O7OET0  
136 144 1.6 1358 2 O7XHO5  
137 144 1.6 1358 2 O8W2P4  
138 144 1.6 1407 2 O9WHY0  
139 144 1.6 1600 2 O7S834  
140 144 1.6 1742 2 O9W350  
141 144 1.6 2693 2 O91908  
142 143.5 1.6 611 2 O9D6C1  
143 143.5 1.6 876 2 O8NR86  
144 143.5 1.6 876 2 C8P19871  
145 143.5 1.6 1287 2 O7QXPI  
146 143.5 1.6 2609 2 O7SG70  
147 143.5 1.6 2609 2 C8E76391  
148 143 1.6 2549 1 FRAP\_HUMAN  
149 143 1.6 2564 2 O7YZ64  
150 143 1.6 5162 2 O701M6

## ALIGNMENTS

## RESULT 1

ID Q14675 PRELIMINARY; PRT; 1745 AA.  
AC Q14675;  
DT 01-NOV-1996 (TRENBLREL. 01. Created)  
DT 01-NOV-1996 (TRENBLREL. 01. Last sequence update)  
DT 01-MAR-2004 (TRENBLREL. 26. Last annotation update)  
DE KIAA0169 protein (Fragment).  
GN Name=KIAA0169;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96281124; PubMed=8724849;  
RA Nagase T., Seki N., Ishikawa K., Tanaka A., Nomura N.;  
RT "Prediction of the coding sequences of unidentified human genes. V.  
RT The coding sequences of 40 new genes (KIAA0161-KIAA0200) deduced by  
RT analysis of cDNA clones from human cell line KG-1.";  
RL DNA Res. 3:17-24(1996).  
RN [2]

RP SEQUENCE FROM N.A.  
RA Ohata O., Nagase T., Kikuno R., Nomura N.,  
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.  
DR EMBL; D79991; BAI1486.1; -  
FT Gene; HGNC:17859; NUP188.  
FT NON TER 1  
SQ SEQUENCE 1745 AA; 195697 MW; CF74DB49C2F6B871 CRC64;  
Query Match 99.2%; Score 8939; DB 2; Length 1745;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1739; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 15 RSSRELTLLIGRALNELSGTAEKMKRRLLEGSTYTPSPSSAEKTKANKDVASP 74  
DB 7 RSSRELTLLIGRALNELSGTAEKMKRRLLEGSTYTPSPSSAEKTKANKDVASP 66  
QY 75 LKEGLNISKPLGIDBESVOLQCYLOEDYGRGDSVKVTLQDERSQALLIKIADYY 134  
DB 67 LKEGLNISKPLGIDBESVOLQCYLOEDYGRGDSVKVTLQDERSQALLIKIADYY 126  
QY 135 EERTCILRCVLLHTYFQDERHPYRVEYACVDKLEKELVSKXROFBEKYTEAPWET 194  
DB 127 EERTCILRCVLLHTYFQDERHPYRVEYACVDKLEKELVSKXROFBEKYTEAPWET 186  
QY 195 HGNMTERQVSRMFWOCLREOSMLELIFYAYFEMAPSDLVLTOMFKOFGSROT 254  
DB 187 HGNMTERQVSRMFWOCLREOSMLELIFYAYFEMAPSDLVLTOMFKOFGSROT 246  
QY 255 RHLVDETMDFVDRIGFYSALLIVEGMDISLHKCALDDREHLPQADGLICODMCLM 314  
DB 247 RHLVDETMDFVDRIGFYSALLIVEGMDISLHKCALDDREHLPQADGLICODMCLM 306  
QY 315 LTFEGDIPHAAPVLLAWALLHRTINPEBTSVVRKIGGTALQLVFYLTLLQSLAGGN 374  
DB 307 LTFEGDIPHAAPVLLAWALLHRTINPEBTSVVRKIGGTALQLVFYLTLLQSLAGGN 366  
QY 375 DCTTSTACMCVYGLSVLTSLSLHTAGNOODITDRCVTLADPSLELWGTGPTSGLG 434  
DB 367 DCTTSTACMCVYGLSVLTSLSLHTAGNOODITDRCVTLADPSLELWGTGPTSGLG 426  
QY 435 IIDSVCMPFPHLSPLQLLRALVSGSKTAKKYSPFLDMSFNNELYKHKPHDIVSHD 494  
DB 427 IIDSVCMPFPHLSPLQLLRALVSGSKTAKKYSPFLDMSFNNELYKHKPHDIVSHD 486  
QY 495 GTLMRRQTPKLLVPLGGQTNLRIPQGTVGQVMDLDRAYLVWMEYSYSWTLFTCEIBML 554  
DB 487 GTLMRRQTPKLLVPLGGQTNLRIPQGTVGQVMDLDRAYLVWMEYSYSWTLFTCEIBML 546  
QY 555 HVTSTADVIOHCRQVPIIDLVHKVSTDSIADCLPITSRIYMLQRLTYISPPVDV 614  
DB 547 HVTSTADVIOHCRQVPIIDLVHKVSTDSIADCLPITSRIYMLQRLTYISPPVDV 606  
QY 615 IASCVNCLVLAARNPAKVTDLRHTGFLPFVAHPVSSLSQMSABGMNAGYGNLMS 674  
DB 607 IASCVNCLVLAARNPAKVTDLRHTGFLPFVAHPVSSLSQMSABGMNAGYGNLMS 666  
QY 675 EOPQSGEVGTIAFLRIITLVKQIGSTOSQGLVPCVMFVLKXMLEPSYHWRVNSHGR 734  
DB 667 EOPQSGEVGTIAFLRIITLVKQIGSTOSQGLVPCVMFVLKXMLEPSYHWRVNSHGR 726  
QY 735 QIGCLILIELHALINCHETDLHSSHTPSLOFCTISLATYBAGQVYINMGVDTIDM 794  
DB 727 QIGCLILIELHALINCHETDLHSSHTPSLOFCTISLATYBAGQVYINMGVDTIDM 786  
QY 795 VMAAPRSDAGSGOGGOLLIKTVKLAFASTVNNVIRLKPSSNVSPLEALSOHAGGN 854  
DB 787 VMAAPRSDAGSGOGGOLLIKTVKLAFASTVNNVIRLKPSSNVSPLEALSOHAGGN 846  
QY 855 LIAVLAKYTHKIDPALPRILAQLRLATVAPMSVYACIGNDAAIRDAFLTRLQSKIB 914  
DB 847 LIAVLAKYTHKIDPALPRILAQLRLATVAPMSVYACIGNDAAIRDAFLTRLQSKIB 906



```

QY 915 DMRIKWMLEFLTAVAVETOPGLIBLFLNLBYKDGSDGSKFSLGWSGLAHVLEIDSDQ 974
Db 907 DMRIKWMLEFLTAVAVETOPGLIBLFLNLBYKDGSDGSKFSLGWSGLAHVLEIDSDQ 966
QY 975 QDRYWCPEPLHRAAIAFLHALMODRDSAMLVLRKPKFWMNTLSPLEFGLSPSSETSEP 1034
Db 967 QDRYWCPEPLHRAAIAFLHALMODRDSAMLVLRKPKFWMNTLSPLEFGLSPSSETSEP 1026
QY 1035 SILETCALIMKICLEIETVYVYKSGIDSLKOTLKKEFSIEKRPAYSGVYKSLAVHVAETE 1094
Db 1027 SILETCALIMKICLEIETVYVYKSGIDSLKOTLKKEFSIEKRPAYSGVYKSLAVHVAETE 1086
QY 1095 GSSCTSLLEYOMLVASRMMLIIATTHADIMHLDSDVYRROLPLDVLDTGATALLVPASV 1154
Db 1087 GSSCTSLLEYOMLVASRMMLIIATTHADIMHLDSDVYRROLPLDVLDTGATALLVPASV 1146
QY 1155 NCURLGSMKCTLLIILRQMRBELGSVDEILGPLETEILEGVLOADQOLMEKTKAVPSAF 1214
Db 1147 NCURLGSMKCTLLIILRQMRBELGSVDEILGPLETEILEGVLOADQOLMEKTKAVPSAF 1206
QY 1215 ITVLQMKEMKVSIDIPOISQVLVNCETIQOEVIYALPDQTRHSLALGSATEDKDSMETDDC 1274
Db 1207 ITVLQMKEMKVSIDIPOISQVLVNCETIQOEVIYALPDQTRHSLALGSATEDKDSMETDDC 1266
QY 1275 SRSRHROROGCVLGLHAKELCEVDSDSMLOVRRLPILPTLLTLEVSILRMKNL 1334
Db 1267 SRSRHROROGCVLGLHAKELCEVDSDSMLOVRRLPILPTLLTLEVSILRMKNL 1326
QY 1335 HFEATLHLLLTLLARTOOGATAVAGAGITQSIGCLPLSVYOLSTNGTQTPSASRKSIDA 1394
Db 1327 HFEATLHLLLTLLARTOOGATAVAGAGITQSIGCLPLSVYOLSTNGTQTPSASRKSIDA 1386
QY 1395 PSMGVYRLSWSLMEQLIKTLRYNPLPEALDFVGHQERTLOCLNAVTVOSLACLEBAD 1454
Db 1387 PSMGVYRLSWSLMEQLIKTLRYNPLPEALDFVGHQERTLOCLNAVTVOSLACLEBAD 1446
QY 1455 HTVGFIIQLSNFMKEMHFLPOLMRDIOVNLGYLCOACTSLHSHRKMLOHLOKNGDGL 1514
Db 1447 HTVGFIIQLSNFMKEMHFLPOLMRDIOVNLGYLCOACTSLHSHRKMLOHLOKNGDGL 1506
QY 1515 PSAVAORVORPPSAASAAPSSSKOPADTEASEOALHTVOYGLKILSKTLAALRHPT 1574
Db 1507 PSAVAORVORPPSAASAAPSSSKOPADTEASEOALHTVOYGLKILSKTLAALRHPT 1566
QY 1575 DVCQILLDOSIDLAENYFLPALSTFTPTFDESEVAPSEFTLLATVVALNMLGELDKKEP 1634
Db 1567 DVCQILLDOSIDLAENYFLPALSTFTPTFDESEVAPSEFTLLATVVALNMLGELDKKEP 1626
QY 1635 LTOAVGLSTOABGRTLKSLMFTMENC FYLLISQANRYLDPVHPDQKRMQELSSSE 1694
Db 1627 LTOAVGLSTOABGRTLKSLMFTMENC FYLLISQANRYLDPVHPDQKRMQELSSSE 1686
QY 1695 LSTLSSLSRYFRGAPSPATGVLPSPQGSTLSKSPSOSPELIDLOVAFYRHMR 1753
Db 1687 LSTLSSLSRYFRGAPSPATGVLPSPQGSTLSKSPSOSPELIDLOVAFYRHMR 1745

```

## RESULT 2

```

06ZOH8
ID 06ZOH8 PRELIMINARY; PRT; 1761 AA.
AC 06ZOH8;
DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DE 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
DE MKIAA0169 protein (fragment).
GN Name=MKIAA0169;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryonic tail;

```

```

RX PubMed=14621295;
RA Okazaki N., Kikuno R., Ohara R., Inamoto S., Koseki H., Hirooka S.,
RA Saga Y., Nagase T., Ohara O., Koga H.;
RT "Prediction of the coding sequences of mouse homologues of KIAA gene:
RT iii. the complete nucleotide sequences of 500 mouse KIAA-homologous
RT cDNAs identified by screening of terminal sequences of cDNA clones
RT randomly sampled from size-fractionated libraries.";
RL DNA Res. 10:167-180(2003).
DR EMBL; AK129073; BAC97883.1; -.
DR InterPro; IPR008938; ARM.
FT NON TER 1
SQ SEQUENCE 1761 AA; 196893 MW; 20CE3DC250EBED93 CRC64;

Query Match 94.2%; Score 8482; DB 2; Length 1761;
Beet Local Similarity 93.9%; Pred. No. 0;
Matches 1643; Conservative 51; Mismatches 45; Indels 10; Gaps 1;

QY 15 RSSRELWTILGSRALRELQIBALNKHMRRLLEGLSYKPPSPSSAEKYANKDVASP 74
Db 13 RSSRELWTILGSRALRELQIBALNKHMRRLLEGLSYKPPSPSSAEKYANKDVASP 72
QY 75 LKELGLRISKPLGDBESQSVQLQCYIQEDYGRTRDSVKTVLQDEROSQALLIKIADYY 134
Db 73 LKELGLRISKPLGDBESQSVQLQCYIQEDYGRTRDSVKTVLQDEROSQALLIKIADYY 132
QY 135 BERTCIIACVILHTTYQDERHPRVAVYADCVDLKEKELSKRQOPEBELYKTEAPMET 194
Db 133 BERTCIIACVILHTTYQDERHPRVAVYADCVDLKEKELSKRQOPEBELYKTEAPMET 192
QY 195 HGNLATEROVSRYMFOCLREOSMLLEIIFLYAAYFENAPSDLVLTWMEKQSGRSQTN 254
Db 193 HGNLATEROVSRYMFOCLREOSMLLEIIFLYAAYFENAPSDLVLTWMEKQSGRSQTN 252
QY 255 RHLVDETMDFPVDRIGFYSALLVVBGMDISLHKCALDDBREHLQPAQDGLICQMDCLM 314
Db 253 RHLVGTMDPVDRIIGFYSALLVVBGMDISLHKCALDDBREHLQPAQDGLICQMDCLM 312
QY 315 LTFGDIYHHAAPVLAAMLLKHTLNPEBTSYVYKIGGTAIQANFYOTLTLQSLASGN 374
Db 313 LTFGDIYHHAAPVLAAMLLKHTLNPEBTSYVYKIGGTAIQANFYOTLTLQSLASGN 372
QY 375 DCTTSTACMCVYGLSLVLSLHTLGNODDIDTACVYADPSLPFLFMGTPTSGIG 434
Db 373 DCTTSTACMCVYGLSLVLSLHTLGNODDIDTACVYADPSLPFLFMGTPTSGIG 432
QY 435 IILDVOCMPHLLSPILLQILRALVSGSKTAKVYSFLDKNSFYNELYKHPHDVISHED 494
Db 433 IILDVOCMPHLLSPILLQILRALVSGSKTAKVYSFLDKNSFYNELYKHPHDVISHED 492
QY 495 GTLMRRQTPKLLYPLGGQTNLRIPQGTGVGYMLDDBRAYLYRWMEYSYSWTLFTCEIEML 554
Db 493 GTLMRRQTPKLLYPLGGQTNLRIPQGTGVGYMLDDBRAYLYRWMEYSYSWTLFTCEIEML 552
QY 555 HVYSTADVIOHCORVKRPIIDLVRKVISIDLSIDCLLPITSRIYMLQRLTTVISPPVDY 614
Db 553 HVYSTADVIOHCORVKRPIIDLVRKVISIDLSIDCLLPITSRIYMLQRLTTVISPPVDY 612
QY 615 IASCNCLTVLAANPAKWTDLRHTGFLPFVAHPVSSLOMISABEGNAGYGNLIMS 674
Db 613 IASCNCLTVLAANPAKWTDLRHTGFLPFVAHPVSSLOMISABEGNAGYGNLIMS 672
QY 675 EOPQGEYGVITIAFLRITTLVYKQSLGSGQVPCVFWLKEMLPBYHKRRYNSHGYRE 734
Db 673 EOPQGEYGVITIAFLRITTLVYKQSLGSGQVPCVFWLKEMLPBYHKRRYNSHGYRE 732
QY 735 QIGCLLIELHALINLCHETDILSSHTPSLOFTICGLATYBAGQVYINIGIGVDITDM 794
Db 733 LIGCLLIELHALINLCOEFLHSSHTPSLCTCGLATYBAGQVYISIGIGVDITDM 792
QY 795 VMAAQRPSDAGGQGOQLIKYKLAFAVYTNVYIRLKPSPNVVSPLEOALSQHGAGNN 854
Db 793 VMAAQRPSDABEGGQGOQLIKYKLAFAVYTNVYIRLKPSPNVVSPLEOALSQHGAGNN 852

```

```

QY 855 LIAVLAKYIYHGDPAIPRLATOLKRLATVA PMSVYACIGNDAAIRDPILRLQSKTE 914
DB 853 LIAVLAKYIYHGDPAIPRLATOLKRLATVA PMSVYACIGNDAAIRDPILRLQSKTE 912
QY 915 DMRIRKMIIEPLTVA VETOPGLIEPLNLEVDGSGSKESFSLGMSCLHVALELIDSOQ 974
DB 913 DMRIRKMIIEPLTVA VETOPGLIEPLNLEVDGSGSKESFSLGMSCLHVALELIDSOQ 972
QY 975 QDRYVCPLLRHAAALAFALHMODRDSAMLVLRTPKRWENUTSPFLFTLSPSETSP 1034
DB 973 QDRYVCPLLRHAAALAFALHMODRDSAMLVLRTPKRWENUTSPFLFTLSPSETSP 1032
QY 1035 SLLETICALMIKICIEIYVVGASLDOSKOTLKFKSIRKPRAYSGYKSLAVVAERE 1094
DB 1033 SVLEFCALIMKICIEIYVVGASLDOSKOTLKFKSIRKPRAYSGYKSLAVVAERE 1092
QY 1095 GSSCTSLLEYOMLVASAMRMLIIATTHADIMHITDSVVRQDFLDVDTKALLVPAV 1154
DB 1093 GSSCTSLLEYOMLVASAMRMLIIATTHADIMHITDSVVRQDFLDVDTKALLVPAV 1152
QY 1155 NCLRIGSKKCTLLILRLQMKELGSDVEIIGPLTEILEGVQADQOLMEKTKAVFSAF 1214
DB 1153 NCLRIGSKKCTLLILRLQMKELGSDVEIIGPLTEILEGVQADQOLMEKTKAVFSAF 1212
QY 1215 ITVLQMKEMKYSIDIPQYSOLVANCETLOEVIYALFDOTRHSIALGSATEDQSMETDC 1274
DB 1213 ITVLQMKELRVQDIPOYSOLVANCETLOEVIYALFDOTRHSIALGSATEDQSMETDC 1272
QY 1275 SRSRRRDQDGVGVGLHAKELCEVDBDGSMLQVTRRLPILPTLITLLEVSLEMKONT 1334
DB 1273 PRPRHKDQDGVGVGLHAKELCEVDBDGSMLQVTRRLPILPTLITLLEVSLEMKONT 1332
QY 1335 HPTREATLHLTLTAATQOGATVAVAGITQSI CIPLLSVYQUSSTNGTQOTPSASKSLDA 1394
DB 1333 HPTREATLHLTLTAATQOGATVAVAGITQSI CIPLLSVYQUSSTNGTQOTPSASKSLDA 1392
QY 1395 PSWPGVYVLSMSLEQLKTLRYNPLPALDFGVHQBRTLOCLNAVTVOSLACLEARD 1454
DB 1393 PSWPGVYVLSMSLEQLKTLRYNPLPALDFGVHQBRTLOCLNAVTVOSLACLEARD 1452
QY 1455 HTVGPIILQLSNPKEMHFLPLQMLMDIQVNLGYLCQACTSLHSHKMLQHTLQNNQDGL 1514
DB 1453 HTVGPIILQLSNPKEMHFLPLQMLMDIQVNLGYLCQACTSLHSHKMLQHTLQNNQDGL 1512
QY 1515 PSAVAQVRQRPESA-----ASAPSSSKOPAAATTEASBOQALHTVGYGLKILSK 1564
DB 1513 PSAVTPRAQRPSTTTTTTTTATATPAGCSSKOPADTEASEOQALHTVGYGLKILSK 1572
QY 1565 TLALRHPTPVCQIILDQSLDLAEVNFALISFTTPTFDESEVAPSFGLLATVVALNM 1624
DB 1573 TLALRHPTPVCQIILDQSLDLAEVNFALISFTTPTFDESEVAPSFGLLATVVALNM 1632
QY 1625 LGEIDKKEPELTOAVGUSTOABGRTTLKSLMFTMENCFFYLLISQAMRYLDPVHPRDK 1664
DB 1633 LGEIDKKEPELTOAVGUSTOABGRTTLKSLMFTMENCFFYLLISQAMRYLDPVHPRDK 1692
QY 1685 QRMKQELSSSELSTLSSLSRFRBGPASBPATGVLPSPQGSTSLSKSPESOBPLQV 1744
DB 1693 QRMKQELSSSELSTLSSLSRFRBGPASBPATGVLPSPQGSTSLSKSPESOBPLQV 1752
QY 1745 QAFVHMQR 1753
DB 1753 QAFVHMQR 1761

```

```

RESULT 3
BAC97883 PRELIMINARY; PRT, 1761 AA.
AC BAC97883;
DT 02-MAR-2004 (TrEMBLrel. 27, Created)
DT 02-MAR-2004 (TrEMBLrel. 27, Last sequence update)
DT 02-MAR-2004 (TrEMBLrel. 27, Last annotation update)
DE KIAA0169 protein (Fragment).

```

```

GN KIAA0169.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISUB=Embryonic tail;
RA Okazaki N., Kikuno R., Ohara R., Inamoto S., Koseki H., Hirooka S.,
RA Suga Y., Nagase T., Ohara O., Koga H.;
RT "Prediction of the Coding Sequences of Mouse Homologues of KIAA Gene:
RT III. The Complete Nucleotide Sequences of 500 Mouse KIAA-homologue
RT cDNAs Identified by Screening of Terminal Sequences of cDNA Clones
RT Randomly Sampled from Size-fractionated Libraries.";
RL DNA Res. 10:167-180(2003).
DR EMBL; AK129073; BAC97883.1; -.
FT NON TER
SQ SEQUENCE 1761 AA; 196893 MW; 20CB3DC250EBBD93 CRC64;

Query Match 94.2%; Score 8482; DB 2; Length 1761;
Best Local Similarity 93.98; Pred. No. 0;
Matches 1643; Conservative 51; Mismatches 45; Indels 10; Gaps 1;

QY 15 RSSRELMTIILGRGALRELQIBALNKHMRRLIEGLSYKPPSPSSAKRYKANKDVASP 74
DB 13 RSSRELMTIILGRGALRELQIBALNKHMRRLIEGLSYKPPSPSSAKRYKANKDVASP 72
QY 75 LKEIGLRISKEPLGDBEBSVOLQCYLOEDYGRTRDSKVTYLDQEROSQALILKTIADYY 134
DB 73 LKEIGLRISKEPLGDBEBSVOLQCYLOEDYGRTRDSKVTYLDQEROSQALILKTIADYY 132
QY 135 EERTCIILCVLHILTYODERHPRVAVYADCYDLKELYSKTRQOEBELYTEAPWET 194
DB 133 EERTCIILCVLHILTYODERHPRVAVYADCYDLKELYSKTRQOEBELYTEAPWET 192
QY 195 HGNLMTROYSRWVQCLREOSMLBIIIFLYAYFENAPSDLVLTLPFKQSGSGRTN 254
DB 193 HGNLMTROYSRWVQCLREOSMLBIIIFLYAYFENAPSDLVLTLPFKQSGSGRTN 252
QY 255 RHIVDETMDPVDRIQFSALILVEGMDIESLHKCALDDBREILHQAQDGLICQMDCLM 314
DB 253 RHIVDETMDPVDRIQFSALILVEGMDIESLHKCALDDBREILHQAQDGLICQMDCLM 312
QY 315 LTRGDIPIHAPVLLAMALLHRTLNPERTSVVRKIGTALQANVQYLTTLLOSLAGCN 374
DB 313 LTRGDIPIHAPVLLAMALLHRTLNPERTSVVRKIGTALQANVQYLTTLLOSLAGCN 372
QY 375 DCTTSTACMCVYGLISVLTSLHTLGNQODIIDTACEVILADPSLPFLPWGTEPTSGLG 434
DB 373 DCTTSTACMCVYGLISVLTSLHTLGNQODIIDTACEVILADPSLPFLPWGTEPTSGLG 432
QY 435 IILDVCGMFPHLISPLQLLRALVSGKSTAKKVSFLDMSFPNELYKKRPHDVISHED 494
DB 433 IILDVCGMFPHLISPLQLLRALVSGKSTAKKVSFLDMSFPNELYKKRPHDVISHED 492
QY 495 GTLMRQTPKLLPLGGQNTLRIPQGTGQVMDLDRAYLVWRWYSVSSWTLPTCEIEMLL 554
DB 493 GTLMRQTPKLLPLGGQNTLRIPQGTGQVMDLDRAYLVWRWYSVSSWTLPTCEIEMLL 552
QY 555 HVTSTADVIOHCORVPIIDLVKRVISTDLSIADCLLPTISRIYMLQRLTTYVSPVDV 614
DB 553 HVTSTADVIOHCORVPIIDLVKRVISTDLSIADCLLPTISRIYMLQRLTTYVSPVDV 612
QY 615 IASCVNCLTVLAARNPAKVTWDLRHGTPLPFVAHPVSSLSQMSIABGMNAGYGNILMS 674
DB 613 IASCVNCLTVLAARNPAKVTWDLRHGTPLPFVAHPVSSLSQMSIABGMNAGYGNILMS 672
QY 675 EOPQGEYGTIAFLRILTYLVQGLSTOSQGLVPCWFTLKMLSYNHRNYSRGVRE 734
DB 673 EOPQGEYGTIAFLRILTYLVQGLSTOSQGLVPCWFTLKMLSYNHRNYSRGVRE 732
QY 735 QIGCLILELHAINTCHETDLSSHTPSIQFICISLAVTEAGQTYINIMIGVPTIDM 794
DB 733 QIGCLILELHAINTCHETDLSSHTPSIQFICISLAVTEAGQTYINIMIGVPTIDM 792

```

Db	733	LIGCILELHIALINLQCETELSHSHTSPSLCISLAIYTBAGTVISINGIDVDTIDM	792
Qy	795	VMAAPRSDGABGCGOGULLITTKVLAPSVTNNVTRILKRPSPNVUSPLEQALSORGAGBN	854
Db	793	VMAAPRSDGEGCGOGULLIKTVLAPSVTNNVTRILKRPSPNVUSPLEQALTOHAGBN	852
Qy	855	LIAVLAKIYHKBHDPRLAIQILKRLATVAPMNVYACIGNDAARIDAFITRLOSKE	914
Db	853	LIAVLAKIYHKBHDPRLAIQILKRLATVAPMNVYACIGSDAAIRIDAFITRLOSKE	912
Qy	915	DMRIKVMILFEFVAVETOPGLIELFLNLVNDGSDGSEFELGMMSCILHAYLELDSOQ	974
Db	913	DMRIKVMILFEFVAVETOPGLIELFLNLVNDGSDGSEFELGMMSCILHAYLELDSOQ	972
Qy	975	QDRYWCPELLHRAAIAFLHALMODRRDSAMVLATKPKFWENTISPLFGTUSPSETSEP	1034
Db	973	QDRYWCPELLHRAAIAFLHALMODRRDSAMVLATKPKFWENTISPLFGTUSPSETSEP	1032
Qy	1035	SILEFICALIMKICIEIYVYVKGSLDOSLKOYLKKFSEKRPAYSGVYKSLAVNAETE	1098
Db	1033	SVLEFICALIMKICIEIYVYVKGSLDOSLKOYLKKFSEKRPAYSGVYKSLAVNAETE	1096
Qy	1095	GSSCSTSLIEYQMLVAMBMILLIATTHADIMELTDSVVRQOLFVLVDGTKALLVLPASV	1155
Db	1093	GSSCSTSLIEYQMLVAMBMILLIATSHADIMELTDMARQOLFVLVDGTKALLVLPASV	1153
Qy	1155	NCLRLGSKCTLLILLRLQWKRELGSVDEILGAPTEILLEGVLQADQOLMEKTKAPPSAF	1214
Db	1153	NCLRLGSKWCTLLILLRLQWKRELGAVERITGLAPTEILLEGVLQADQOLMEKTKAPPSAF	1212
Qy	1215	ITVLQMKEMKASDIPOYSOLVLANVETLOEBVIALPDQTRHSLAGSATBEDSDMETDC	1274
Db	1213	ITVLQMKELRAGDIPQYSOLVLANVETLOEBVIALPDQTRHSLADSAEBDSMETDC	1272
Qy	1275	SRSRRDRDQGVCGVLGHLAKELCEVEDGDSMLQVTRSLPLITLITLIEVLSMKONTL	1334
Db	1273	PRPRKQDRDQGVCGVLGHLAKELCEVEDGDSMLQVTRSLPLITLITLIEVLSMKONTL	1332
Qy	1335	HFTBATLHLLTLAFTQOGATVAVAGITQOSICPLLSVYQLSNGTNGAOTPSAKSKSIDA	1394
Db	1333	HFTBATLHLLTLAFTQOGATVAVAGITQOSICPLLSVYQLSNGTNGAOTPSAKSKSIDA	1392
Qy	1395	PSMPGVYRLSNSLMEOULKTLRYNPLPEALDVGVGHQERTLOCNANVTQSLACLERAD	1454
Db	1393	PSMPGVYRLSNSLMERLLKTLRYNPLEALDVGVGHQERTLOCNANVTQSLACLERAD	1452
Qy	1455	HTVGTIILQLSNMKEMHFLPOLMADIONVLGYLOACSTSLHSKMLQHYLQNNKGDL	1514
Db	1453	HTVGTIILQLSHRKEMHFLPOLMADIONVLGYLOACSTSLHSKMLQHYLQNNKGDL	1512
Qy	1515	PSAVALQVORPSPA-----ASAPSSSKOPAADTEASEQOALHTVOYGLKITLSK	1564
Db	1513	PSAVALQVORPSTTTTTTTTTLNALTNPACSSKOPAADTEASEQALHTVOYGLKITLSR	1562
Qy	1565	TLAALRHFTPDVCOILLDQSLDLAEYNLFALSFTTPTFDSVADSGFTLATVNAVLM	1622
Db	1573	TLAALRHFTPDVCOILLDQSLDLAEYNLFALSFTTPTFDSVADSGFTLATVNAVLM	1620
Qy	1635	IGELDKKKEPRLTOANGVGLSTQOAGSTRLLKSLMFTMENCYLLISQAMRVLRAVAHPRDX	1684
Db	1633	IGELDKKKESTLQAVGLSTQOAGSTRLLKSLMFTMENCYLLISQAMRVLRAVAHPRK	1682
Qy	1685	QRMKOBLSSELSTLSSLSRYPFRGAPSPAPGVLPSPGKSTSLSKAPSEQBPILQV	1744
Db	1683	QRMKOBLSSELSTLSSLSRYPFRGAPSPAPGVLPSPGKSTSLSKAPSEQBPILQV	1742
Qy	1745	QAFVHRMOR 1753	
Db	1753	QAFVHRVQR 1761	
RESULT 4			
Q7ZJX8			

ID	Q723K8	PRELIMINARY;	PR: 1639 AA.
AC	Q723K8		
DT	01-OCT-2003 (TrEMBLrel. 25, Created)		
DT	01-OCT-2003 (TrEMBLrel. 25, Last sequence update)		
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)		
GN	Hypothetical protein DKFZp686l1653 (Fragment).		
GN	Name=DKFZp686l1653;		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.		
OX	NCBI_Taxid=9606;		
BN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=Human endometrium;		
RA	Wambut R., Heubner D., Mewes H.W., Weil B., Amid C., Osanger A.,		
RA	Fobo G., Han M., Mlemane S.;		
RL	Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; BX537774; CAD97835.1; -		
KM	Hypothetical protein.		
FT	NON TER		
SO	SEQUENCE 1639 AA; 182478 MW; B58334E14B217B92 CRC64;		
	Query Match 91.8%; Score 8270.5; DB 2; Length 1639;		
	Best Local Similarity 93.5%; Pred. No. 0;		
	Matches 1626; Conservative 1; Mismatches 1; Indels 11; Gaps 1		
QY	15 RSSRLWTLILGRSALRELSCIEALNKGWRLLEGLSYKPPSPSSAEKVKANDVASP	74	
DB	12 RSSRLWTLILGRSALRELSCIEALNKGWRLLEGLSYKPPSPSSAEKVKANDVASP	71	
QY	75 LKEIGLRISKEGLDEBOSVOLLCYLOEDVGRGDSVKTYLQDNRQSGALLIKLADYY	134	
DB	72 LKEIGLRISKEGLDEBOSVOLLCYLOEDVGRGDSVKTYLQD-----	115	
QY	135 EERTCILRCVLAHLTTYFQDERHPYKRVAYADCVDKLEKELVSKYRQOFBELYKTEAPWTET	194	
DB	116 -----	115	
QY	195 HGNLWTERQVSRMFWQCRLREQSMLEITFLYYAFEMAPSDLLVLTGKFKEGFGSGRTN	254	
DB	116 -----EMAPSDLLVLTGKFKEGFGSGRTN	140	
QY	255 RHLVDETMDPVDRIGYRSALILVNGMIEISLHKCALDRRLHOPADQGLCOMDCLM	314	
DB	141 RHLVDETMDPVDRIGYRSALILVNGMIEISLHKCALDRRLHOPADQGLCOMDCLM	200	
QY	315 LTFGDIPIHAAPVLAAMALLRHTLNPEETSSVVRKIGTATQIANVQYLTRLQSLASG	374	
DB	201 LTFGDIPIHAAPVLAAMALLRHTLNPEETSSVVRKIGTATQIANVQYLTRLQSLASG	260	
QY	375 DQTSSTACMCYVGLISFVLTSLRLATTGNNQDIDTACRVLADPSLPFLFWGTETPTSGIG	434	
DB	261 DQTSSTACMCYVGLISFVLTSLRLATTGNNQDIDTACRVLADPSLPFLFWGTETPTSGIG	320	
QY	435 IILDSVCGMPHLLSPILQLRALVSGSTAKKYSPLDKMSFNNELVYGRKHDIYSHD	494	
DB	321 IILDSVCGMPHLLSPILQLRALVSGSTAKKYSPLDKMSFNNELVYGRKHDIYSHD	380	
QY	495 GLWNRQTPKLLLYPLGGQTNLRIPQGTQVQWLDRAVLYVREYSSWTFTCEIEMLL	554	
DB	361 GLWNRQTPKLLLYPLGGQTNLRIPQGTQVQWLDRAVLYVREYSSWTFTCEIEMLL	440	
QY	555 HVSSTADVIQHCQRVKPIIDLVHAKVISITDLSIADCLPITSRIYMLQRLTVVISPPDV	614	
DB	441 HVSSTADVIQHCQRVKPIIDLVHAKVISITDLSIADCLPITSRIYMLQRLTVVISPPDV	500	
QY	615 IASCNCLTVLAARPAKVTMDLRHTGLPFVAHVSSLSQMSIABGNAGGYNLNN	674	
DB	501 IASCNCLTVLAARPAKVTMDLRHTGLPFVAHVSSLSQMSIABGNAGGYNLNN	560	
QY	675 EOPQGEYGVTAFLALITLTVKQGLSTQSGQVPCVNFVLEMLPSYKRWIRNSHGVSE	734	
DB	561 EOPQGEYGVTAFLALITLTVKQGLSTQSGQVPCVNFVLEMLPSYKRWIRNSHGVSE	620	

```

QY 735 QIGCLILELTHALINLCHETDLSHSHPSIOFLCISLAATYBAGOTVINMGIGVDTIDM 794
DB 621 QIGCLILSHLHAILNLCHEIDLSHSHPSIOFLCISLAATYBAGOTVINMGIGVDTIDM 680
QY 795 VMAAOPRSDGAEQGGGQGLIKTYKLAFSVTNNVIRLKPSPNVVSPLEQALSQGHAGNN 854
DB 681 VMAAOPRSDGAEQGGGQGLIKTYKLAFSVTNNVIRLKPSPNVVSPLEQALSQGHAGNN 740
QY 855 LIAVLAKYIKHNDPALPRALIQLLKSLATYAPMSVYACIGANDAAIRDAFLTRLOSKIE 914
DB 741 LIAVLAKYIKHNDPALPRALIQLLKSLATYAPMSVYACIGANDAAIRDAFLTRLOSKIE 800
QY 915 DMEIKWILEFLVAVEVETOPGLIELFNLBYKQSDSDSKPSLGMSCLAHVELIDISQO 974
DB 801 DMEIKWILEFLVAVEVETOPGLIELFNLBYKQSDSDSKPSLGMSCLAHVELIDISQO 860
QY 975 QDRYWCPELHRAAIAFLHLMODRDSAMLVIRTKPEWMENTSPULFGTSLSPSETSEP 1034
DB 861 QDRYWCPELHRAAIAFLHLMODRDSAMLVIRTKPEWMENTSPULFGTSLSPSETSEP 920
QY 1035 SILETCALIMKIIICLETYYVVKSGSDSLKDTLKSEIERKPAWSGYVKSIAVHVAET 1094
DB 921 SILETCALIMKIIICLETYYVVKSGSDSLKDTLKSEIERKPAWSGYVKSIAVHVAET 980
QY 1095 GSSCTSLLEYQMLVSARMLIIATTHADIMHLDSDSVVRQOLFVDVDTGKALLVPAV 1154
DB 981 GSSCTSLLEYQMLVSARMLIIATTHADIMHLDSDSVVRQOLFVDVDTGKALLVPAV 1040
QY 1155 NCIRLGSMTCTLLIILIRQKRELSYVDILGLPTEILBEGVLOADQOLMEKTKAYPSAF 1214
DB 1041 NCIRLGSMTCTLLIILIRQKRELSYVDILGLPTEILBEGVLOADQOLMEKTKAYPSAF 1100
QY 1215 ITVLQKEMKENVSDIPOYSQVLNVCETLOREVALPQOTHSIALSGATDCKSMETDCC 1274
DB 1101 ITVLQKEMKENVSDIPOYSQVLNVCETLOREVALPQOTHSIALSGATDCKSMETDCC 1160
QY 1275 SRSRHRDQDGVCLGLHLAKELCEVEDDSDMLQVTRRLPIPTLLITLVESLRKNOL 1334
DB 1161 SRSRHRDQDGVCLGLHLAKELCEVEDDSDMLQVTRRLPIPTLLITLVESLRKNOL 1220
QY 1335 HFTBATLHLLLTARTQOGATAVAGAGITQSIGLPILSYVQSLSTNGTACTPSASRSLDA 1394
DB 1221 HFTBATLHLLLTARTQOGATAVAGAGITQSIGLPILSYVQSLSTNGTACTPSASRSLDA 1280
QY 1395 PSWPGYRLSMSLMEOLKTLRYNPLPEALDFVGHQERTLOCLNAVRTVQSLACLEAD 1454
DB 1281 PSWPGYRLSMSLMEOLKTLRYNPLPEALDFVGHQERTLOCLNAVRTVQSLACLEAD 1340
QY 1455 HTVGFILQSLNFKEMHFLPOLMRDIOVNLGYLCOACTSILASRKNOLQHYLONKQDGL 1514
DB 1341 HTVGFILQSLNFKEMHFLPOLMRDIOVNLGYLCOACTSILASRKNOLQHYLONKQDGL 1400
QY 1515 PSVAVORVORPPSASAPSSSKOPADTASQOALHTVOYGLKILSKTALRHFP 1574
DB 1401 PSVAVORVORPPSASAPSSSKOPADTASQOALHTVOYGLKILSKTALRHFP 1460
QY 1575 DVCQILDDSLDAENFLPALSFPTPTDSEVAFRGTLTAVNALMGLDKKKEP 1634
DB 1461 DVCQILDDSLDAENFLPALSFPTPTDSEVAFRGTLTAVNALMGLDKKKEP 1520
QY 1635 LTQAVGLSTQAGSTRILKSLMFTMENCFFILISQAMRYLDRPAVHPRDKOMKQELSE 1694
DB 1521 LTQAVGLSTQAGSTRILKSLMFTMENCFFILISQAMRYLDRPAVHPRDKOMKQELSE 1580
QY 1695 LSTLLSLSRYRRGAPSSPATGVLPSPQKSTSLSKASPESEGLIQLVQAFVRMOR 1753
DB 1581 LSTLLSLSRYRRGAPSSPATGVLPSPQKSTSLSKASPESEGLIQLVQAFVRMOR 1639

```

RESULT 5  
Q81WF1 PRELIMINARY; PRT; 1293 AA.

```

AC Q81WF1;
DT 01-MAR-2003 (Tremblere). 23, Created)
DT 01-MAR-2003 (Tremblere). 23, Last sequence update)
DT 01-MAR-2004 (Tremblere). 26, Last annotation update)
DE NUP188 protein (Fragment).
GN Name=NUP188;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Ovary;
RX MEDLINE=22386257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Krausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heien F.,
RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udell T.B., Toshiyuki S., Carninci P., Frange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Millar S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Vallalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Pahay J., Helton E., Kerteman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzyzanski M.I., Skalska U., Smalls D.E., Scherch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Ovary;
RA Strausberg R.;
RA Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC040352; AAK0352.1; -.
DR InterPro; IPR008938; ARM.
FT NON TER
SQ SEQUENCE 1293 AA; 144063 MW; F0D1B7BE69DC1791 CRC64;
Query Match 73.4%; Score 6608; DB 2; Length 1293;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1292; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 461 GKSTAKKYSPFADKSPFYNELYKHPHVDISHEDGLMWRQPKLLYPGQTNLRIPGG 520
DB 1 GKSTAKKYSPFADKSPFYNELYKHPHVDISHEDGLMWRQPKLLYPGQTNLRIPGG 60
QY 521 TVGQVMDLDRAYLVWEYSYSWSWTLFTCEIEKMLHVSSTADVIQHCQRYKPIIDLVHXY 580
DB 61 TVGQVMDLDRAYLVWEYSYSWSWTLFTCEIEKMLHVSSTADVIQHCQRYKPIIDLVHXY 120
QY 581 STDLSIADCLPITSRIYMLQRLTTVISPPDVVASCVNCLTVAAARPAVMDLRHT 640
DB 121 STDLSIADCLPITSRIYMLQRLTTVISPPDVVASCVNCLTVAAARPAVMDLRHT 180
QY 641 GFLPFAHVPVSLSGMISABGNAGYGNLMMNSBQPGBYGTIAPRLITTLVKGQGL 700
DB 181 GFLPFAHVPVSLSGMISABGNAGYGNLMMNSBQPGBYGTIAPRLITTLVKGQGL 240
QY 701 STQSGGLVPCVWFVLEKMLPSYHKRRYNSHGVRQIGCLILELTHALINLCHETDLSHSH 760
DB 241 STQSGGLVPCVWFVLEKMLPSYHKRRYNSHGVRQIGCLILELTHALINLCHETDLSHSH 300
QY 761 TPSLOFLCISLAATYBAGOTVINMGIGVDTIDMMAAOPRSDGAEQGGGQGLIKTYKL 820
DB 301 TPSLOFLCISLAATYBAGOTVINMGIGVDTIDMMAAOPRSDGAEQGGGQGLIKTYKL 360
QY 821 AFSVTNNVIRLKPSPNVVSPLEQALSQGHAGNNLIAVLAKYIKHNDPALPRALIQLLK 880

```

```

Db 361 AFSVTNNVIRLKPENSVSPLEQALSOHAGHGNLIVAKIYKHDPALRLAIQLX 420
Qy 861 RLATVAPMSVYACLGNDAAAIKDAFLTRLOSKIEMRIKVMILBFLVAVETOPGLIEF 940
Db 421 RLATVAPMSVYACLGNDAAAIKDAFLTRLOSKIEMRIKVMILBFLVAVETOPGLIEF 480
Qy 941 LNLVKKDSDGSKESLGMWCLHVLIELIDSOQDRVWCPPLLRALAFHAIAMORR 1000
Db 481 LNLVKKDSDGSKESLGMWCLHVLIELIDSOQDRVWCPPLLRALAFHAIAMORR 540
Qy 1001 DSAMVLRTRKPFENLTSPLFGLTSPSPSESEPSILFETCALIMKICILEIYVYVKSJD 1060
Db 541 DSAMVLRTRKPFENLTSPLFGLTSPSPSESEPSILFETCALIMKICILEIYVYVKSJD 600
Qy 1061 QSLKOTLKKFSIEKRFAYWSGYKSLAVHVAETEGSSCTSLLEYQMLVSAMRMLLIAT 1120
Db 601 QSLKOTLKKFSIEKRFAYWSGYKSLAVHVAETEGSSCTSLLEYQMLVSAMRMLLIAT 660
Qy 1121 HADIMHLDVSVRRQLFDVLDGTAKALLVPASVNCRLRSGMKCTLLILRQMKRELS 1180
Db 661 HADIMHLDVSVRRQLFDVLDGTAKALLVPASVNCRLRSGMKCTLLILRQMKRELS 720
Qy 1181 VDEILGPILEIEGLVQADQOMKTKAKYFSAFIVLQMKMKYSDIPQYQSLVYVNC 1240
Db 721 VDEILGPILEIEGLVQADQOMKTKAKYFSAFIVLQMKMKYSDIPQYQSLVYVNC 780
Qy 1241 TLQEEVIALFDQTRHSLALGSATEDKSDMETDCSRHRRDRDQVGVGLHLAKELCEV 1300
Db 781 TLQEEVIALFDQTRHSLALGSATEDKSDMETDCSRHRRDRDQVGVGLHLAKELCEV 840
Qy 1301 DEDGSMVQVTRRLIPLTLITLLEVSLEMKONLHTEKTLHLTLATQOGATRVAGA 1360
Db 841 DEDGSMVQVTRRLIPLTLITLLEVSLEMKONLHTEKTLHLTLATQOGATRVAGA 900
Qy 1361 GITGSIQPLSVYQSLSTNGTQOTPSASRKSIDAPSPCVRLSINSMEQLKTLRYNPL 1420
Db 901 GITGSIQPLSVYQSLSTNGTQOTPSASRKSIDAPSPCVRLSINSMEQLKTLRYNPL 960
Qy 1421 PEALDFVGHQERTLOCLNAVTVOSLACLEADHVTGFILOJNSFMKWHFHLPLQMD 1480
Db 961 PEALDFVGHQERTLOCLNAVTVOSLACLEADHVTGFILOJNSFMKWHFHLPLQMD 1020
Qy 1481 IQVNLGYLQACTSLHSHKMLQHTIQNKXGDLPSAVQVQRPSPASAAAPSSSKQPA 1540
Db 1021 IQVNLGYLQACTSLHSHKMLQHTIQNKXGDLPSAVQVQRPSPASAAAPSSSKQPA 1080
Qy 1541 ADTEASBOQALHTVOYGLIKITLTAALRHFTPDVCOILLDOSLDLAEYNFLFALSFT 1600
Db 1081 ADTEASBOQALHTVOYGLIKITLTAALRHFTPDVCOILLDOSLDLAEYNFLFALSFT 1140
Qy 1601 PTFDESVAPSGTLLATVAVNALMGLDKKKEPLTOAVGLSTOAGRTLLKSLMTME 1660
Db 1141 PTFDESVAPSGTLLATVAVNALMGLDKKKEPLTOAVGLSTOAGRTLLKSLMTME 1200
Qy 1661 NCFYLLISQAMRYLRDPAVHPDKQPMKQELSSSELSTLSSIRYRRQAPSSPATGVP 1720
Db 1201 NCFYLLISQAMRYLRDPAVHPDKQPMKQELSSSELSTLSSIRYRRQAPSSPATGVP 1260
Qy 1721 SPQGSSTLSKASPSQEPRLIOLVOAFVHHMR 1753
Db 1261 SPQGSSTLSKASPSQEPRLIOLVOAFVHHMR 1293

```

## RESULT 6

```

Q9BS12 PRELIMINARY; PRT; 853 AA.
ID Q9BS12
AC Q9BS12;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE NUP186 protein (fragment).
GN Name=NUP186,

```

```

OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OC NCBI_taxid=9606;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Cervix;
RX MEDLINE=22388257; PubMed=12477932;
RA Struhsberg R.L., Pelngold R.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buelow K.H., Schaefer C.F., Bat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diachenko L., Marushina K., Farmer A.A., Rubin G.W., Hong L.,
RA Stapleton M., Soares M.B., Bonaldi M.F., Carantini T.L., Prange C.,
RA Brownstein M.J., Usdin T.B., Toshynki S., Abramson R.D., Mullaly S.J.,
RA Rane S.S., Loguellano N.A., Peters G.J., Abrahams R.D., Gunaratne P.H.,
RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.W., Butlerfield Y.S.,
RA Krzywinski M.I., Skalski U., Smalios D.E., Schnerch A., Schein J.B.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN (2)
RP SEQUENCE FROM N.A.
RC TISSUE=Cervix;
RA Struhsberg R.;
RL Submitted (MAR-2001) to the EMBL/Genbank/DBJ databases.
DR EMBL; BC005407; AA005407.1; -.
FT NON TER
SQ SEQUENCE 853 AA; 94944 MW; F9BD8397BE7ED004 CRC64;

```

```

Query Match 45.4%; Score 4093; DB 2; Length 853;
Best Local Similarity 96.2%; Pred. No. 7.4e-265;
Matches 810; Conservative 5; Mismatches 21; Indels 6; Gaps 1;
Qy 683 VTIAFLRLITLVKQGLSTQSGVPCMFVLEKMLPSYHMKRYNSGVREIQCLILE 742
Db 1 VTIAFLRLITLVKQGLSTQSGVPCMFVLEKMLPSYHMKRYNSGVREIQCLILE 60
Qy 743 LIHALINTCHETDLSHTSPSIQFLCISLAVTEAGQVYINIMIGIVDTIDVMAAOPRS 802
Db 61 LIHALINTCHETDLSHTSPSIQFLCISLAVTEAGQVYINIMIGIVDTIDVMAAOPRS 120
Qy 803 DGABGQGGQGLIKTVKLAFSYTNVNIIRLKPENSVSPLEQALSOHAGHGNLIVLAKY 862
Db 121 DGABGQGGQGLIKTVKLAFSYTNVNIIRLKPENSVSPLEQALSOHAGHGNLIVLAKY 180
Qy 862 IYHKDPAFLRLAIDLRATVAPMSVYACLGNDAAAIKDAFLTRLOSKIEMRIKVM 922
Db 181 IYHKDPAFLRLAIDLRATVAPMSVYACLGNDAAAIKDAFLTRLOSKIEMRIKVM 240
Qy 922 LEFLVAVETOPGLIEFLNLNLEVKDQSGSKFSLGMSCLHVLIELIDSOQDRVWCP 982
Db 241 LEFLVAVETOPGLIEFLNLNLEVKDQSGSKFSLGMSCLHVLIELIDSOQDRVWCP 300
Qy 982 LIHRAAIAFLHAIWDRRDSANLVIRTRKPFENLTSPLFGLTSPSPSESEPSILFETCAL 1042
Db 301 LIHRAAIAFLHAIWDRRDSANLVIRTRKPFENLTSPLFGLTSPSPSESEPSILFETCAL 360
Qy 1043 IMKICLRIYVYVKSGLDOSLQDLKKFSIEKRFAYWSGYKSLAVHVAETEGSSCTSL 1102
Db 361 IMKICLRIYVYVKSGLDOSLQDLKKFSIEKRFAYWSGYKSLAVHVAETEGSSCTSL 420
Qy 1103 EYQMLVSAMRMLLIATTHADIMHLDVSVRRQLFDVLDGTAKALLVPASVNCRLRSGM 1162
Db 421 EYQMLVSAMRMLLIATTHADIMHLDVSVRRQLFDVLDGTAKALLVPASVNCRLRSGM 480

```



DB 491 GTLROROTPLLYPGGTNRIRPGIVGQWMDRALVWREISYSSWTFTECEHLL 550  
 QY 555 HVASTADVIQHCORVPIIDLVHXYISTDLSIADCLPITSRIYMLQRLTTVISPVDV 614  
 DB 551 HVASTADVIQHCORVPIIDLVHXYISTDLSIADCLPITSRIYMLQRLTTVISPVDV 610  
 QY 615 IASVCNCLTVLAARPAVWTDLRHTGFLPFAHVSLSQMSIEGNNAGYGLTMS 674  
 DB 611 IASVCNCLTVLAARPAVWTDLRHTGFLPFAHVSLSQMSIEGNNAGYGLTMS 670  
 QY 675 EPOGEGVITAFILRLITTLVKGQSGTOSQGLVPCWFLKEMLPYHKKRYNSHGYRE 734  
 DB 671 EPOGEGVITAFILRLITTLVKGQSGTOSQGLVPCWFLKEMLPYHKKRYNSHGYRE 730  
 QY 735 QIGCLIELLHAIINLCHETDLHSHTPSLOFLCISLAYTEAGOTVINIGIVDTIDM 794  
 DB 731 LIGCLIELLHAIINLCHETDLHSHTPSLOFLCISLAYTEAGOTVINIGIVDTIDM 790  
 QY 795 VMAAOPR 801  
 DB 791 VMAAOPR 797  
 RESULT 8  
 QY 08R3F1 PRELIMINARY; PRT; 525 AA.  
 AC 08R3F1  
 DT 01-JUN-2002 (TREMBlrel. 21, Created)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)  
 DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)  
 DE Nup188 protein (fragment).  
 GN Name=BC025526; Synonyms=Nup188;  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxId=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FVB/N; TISSUE=Mammary tumor. C3;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,  
 Altschul S.F., Zeeberg B., Buetow K.H., Scheffer C.F., Bat N.K.,  
 Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullighy S.J.,  
 Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 Fahey J., Helton E., Kerteman M., Madan A., Rodriguez S., Sanchez A.,  
 Whitling M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butlerfield Y.S.,  
 Krzywinski M.I., Skalska U., Smalins D.E., Scherch A., Schein J.B.,  
 Jones S.J., Marra M.A.,  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 and mouse cDNA sequences";  
 RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [12]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FVB/N; TISSUE=Mammary tumor. C3;  
 RA Strauberg R.;  
 RA Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC025526; AAH25526.1;  
 DR MGD; MGI:2446190; BC025526.  
 FT NON\_TER  
 SQ SEQUENCE 525 AA; 58114 MW; D60B0FB81DAE2CE CRC64;  
 Query Match 26.9%; Score 2420; DB 2; Length 525;  
 Best Local Similarity 91.6%; Pred. No. 3.8e-153;

Matches 478; Conservative 13; Mismatches 21; Indels 10; Gaps 1;  
 QY 1242 LOERVIALPOTRSLALGSTRTEKDSEMTDDCSRSRHRDQDVCYGLHAKELCEVD 1301  
 DB 4 LOERVIALPOTRSLALGSTRTEKDSEMTDDCSRSRHRDQDVCYGLHAKELCEVD 63  
 QY 1302 EDGDSMLQVTRRLILPILTLTLEVSILRMKONLHFTTEHTLLTLTLARTQCATAVAG 1361  
 DB 64 EDGDSMLQVTRRLILPILTLTLEVSILRMKONLHFTTEHTLLTLTLARTQCATAVAG 123  
 QY 1362 ITGQICPLLSVYQSLNGTNAQTSPASRKSIDAPSWGVYLSMSLEQLKTRRYNLP 1421  
 DB 124 ITGQICPLLSVYQSLNGTNAQTSPASRKSIDAPSWGVYLSMSLEQLKTRRYNLP 183  
 QY 1422 EALPVGVRHQRITQCLNAKRTVQSLACLEADHTTGGITLQSPMKEMHFLPOLMDV 1481  
 DB 184 EALPVGVRHQRITQCLNAKRTVQSLACLEADHTTGGITLQSPMKEMHFLPOLMDV 243  
 QY 1482 QVNLGYLCOACTSLHSRKMQLHYLQNNKNGDGLPSAVALVQRPSPA-----ASA 1531  
 DB 244 QVNLGYLCOACTSLHSRKMQLHYLQNNKNGDGLPSAVALVQRPSPA-----ASA 303  
 QY 1532 APSSKQPADTBASBOALHTVQYGLKILSKTLAALRHFTPPVQCILLDQSLDAEYN 1591  
 DB 304 AGSSKQPADTBASBOALHTVQYGLKILSKTLAALRHFTPPVQCILLDQSLDAEYN 363  
 QY 1592 FLPLALFTPTPDEEVAPSFETLLATVVALNMGLGDKKEPILTONAGSTOAGRTLL 1651  
 DB 364 FLPLALFTPTPDEEVAPSFETLLATVVALNMGLGDKKEPILTONAGSTOAGRTLL 423  
 QY 1652 KSLMFTWNCFFYLLISQAMRYLRDPAVHPDKRMKQELSESLTLLSISRYRRGAP 1711  
 DB 424 KSLMFTWNCFFYLLISQAMRYLRDPAVHPDKRMKQELSESLTLLSISRYRRGAP 483  
 QY 1712 SSPATGVLPSPQKSTISKASPSQEPILQVQAFVHROR 1753  
 DB 484 SSPATGVLPSPQKSTISKASPSQEPILQVQAFVHROR 525  
 RESULT 9  
 QY 080UL4 PRELIMINARY; PRT; 403 AA.  
 AC 080UL4  
 DT 01-JUN-2003 (TREMBlrel. 24, Created)  
 DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)  
 DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)  
 DE Nup188 protein (fragment).  
 GN Name=BC025526; Synonyms=Nup188;  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxId=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=B6/EGFP transgenic ICR mice; TISSUE=Trophoblast Stem Cell;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,  
 Altschul S.F., Zeeberg B., Buetow K.H., Scheffer C.F., Bat N.K.,  
 Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 Datchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullighy S.J.,  
 Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 Fahey J., Helton E., Kerteman M., Madan A., Rodriguez S., Sanchez A.,  
 Whitling M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butlerfield Y.S.,  
 Krzywinski M.I., Skalska U., Smalins D.E., Scherch A., Schein J.B.,  
 Jones S.J., Marra M.A.,



"Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences.";  
Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

RP SEQUENCE FROM N.A.  
RC STRAIN=BS/EGFP transgenic ICR mice; TISSUE=Trophoblast Stem Cell;

RA Strubeberg R.;  
Submitted (Apr-2003) to the EMBL/GenBank/DBJ databases.

DR EMBL; BC050199; AAH50199.1; -  
DR MGI; MGI:2446190; BC025526.

FT NON TER 1  
SQ SEQUENCE 403 AA; 44577 MW; 7BC6D3C8FEC33802 CRC64;

Query Match 20.6%; Score 1852; DB 2; Length 403;  
Best Local Similarity 90.8%; Pred. No. 2.7e-115;  
Matches 366; Conservative 12; Mismatches 15; Indels 10; Gaps 1;

QY 1361 GTTOSTCLPRLAYOUSTNGTGTPTASRSKSLDAPSPGVYRLSMGLMGLKTLAYNPL 1420  
DB 1 GTTOSTCLPRLAYOUSTNGTGTPTASRSKSLDAPSPGVYRLSMGLMGLKTLAYNPL 60

QY 1421 PALADPVGHOERTLOCLNAVTVQSLACLEADHTVGFLLQTSNFMKSHFLLPOLMRD 1480  
DB 61 TERLADPVGHOERTLOCLNAVTVQSLACLEADHTVGFLLQTSNFMKSHFLLPOLMRD 120

QY 1481 IGVNIGYLCQACTSLHSHRKLQHYLQNKXGDELPSAVAKRVQRPSPA-----AS 1530  
DB 121 VGVNIGYLCQACTSLHSHRKLQHYLQNKXGDELPSAVAKRVQRPSPA-----AS 180

QY 1531 AAPSSKOPADPTESECOALHTVQYGLKTLKTLTAALRHFTPDVCOILLDQSLDARY 1590  
DB 181 PGGSSKOPADPTESECOALHTVQYGLKTLKTLTAALRHFTPDVCOILLDQSLDARY 240

QY 1591 NFLFALFTPTFDSVAPSFGLTLATVVAALNMLGELDKKKKPLTQAVGLSTQAGETRT 1650  
DB 241 NFLFALFTPTFDSVAPSFGLTLATVVAALNMLGELDKKKKPLTQAVGLSTQAGETRT 300

QY 1651 LKSLMFTWENCFFYLLISQAMRYLDPVAHPRDKQKQKQKSELSTLSSLRFFRGA 1710  
DB 301 LKSLMFTWENCFFYLLISQAMRYLDPVAHPRDKQKQKQKSELSTLSSLRFFRGA 360

QY 1711 BSSPAAGVLPSPGKATSLSKASPSQEPILQLVQAFVRRVOR 1753  
DB 361 BSSPAAGVLPSPGKATSLSKASPSQEPILQLVQAFVRRVOR 403

# RESULT 10

QY09K3 PRELIMINARY; PRT; 1056 AA.

AC QY09K3  
DT 01-MAR-2004 (TrEMBLrel. 26, Created)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE AGCP15000 (Fragment).

GN Name=agc46168; ORFName=ENSG00000007942;  
OS Anopheles gambiae str. PEST.  
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.

OC NCBI\_TaxId=180454;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=PEST;

RA Anopheles Genome Sequencing Consortium;  
Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
CC -! CAUTION: The sequence shown here is derived from an  
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
CC preliminary data.

DR EMBL; AAA01008900; BAA09356.1; -  
FT NON TER 1  
FT NON TER 1056  
SQ SEQUENCE 1056 AA; 121746 MW; 191129F749BA27A2 CRC64;

Query Match 9.4%; Score 848; DB 2; Length 1056;

Best Local Similarity 26.2%; Pred. No. 2.1e-47;  
Matches 291; Conservative 207; Mismatches 478; Indels 136; Gaps 39;

QY 18 RELWTILLGRSLRELQIEALNKHWRLLBGLSYKPPSSAKRY-KAKDVAS-PL 75  
DB 3 RELWTILLGRSLRELQIEALNKHWRLLBGLSYKPPSSAKRY-KAKDVAS-PL 62

QY 76 KELGRIKSLGLDDEQSVOLLQCTLODYGTGTRSVTVYLODERQSOALLIKINDYYE 135  
DB 63 QPFAKMLQYDIDVQSWQILCYTLVNEYGAAVLALEYISTEISWVLLHDIWTHYL 122

QY 136 ERTCLRCYLLHLYFODERHPYRYEYADCVKLEKELVSK-YROQFELYKTEAPTW- 193  
DB 123 ERMVQLKMKKLLBETHSGTHPYSKYEVKEMGLQRLKSYLAQTLIL-ADPPSSQK 181

QY 194 -----THGNLMTERQVSRWFVOC---LRQSMLEIIFLYVAFEMAPSDLLVLYM 242  
DB 182 APLAGDLHGGRR-----VACABRLRETNELIQLILLIYVAGISPEELDLQFL 232

QY 243 FREGQFGRQTRHLYVDETNP-----PYDRIGYFSAIIVEGMDISLHKALDDRR 295  
DB 243 FREGQFGRQTRHLYVDETNP-----PYDRIGYFSAIIVEGMDISLHKALDDRR 281

QY 296 ELHOPADQGLICQDNDCLMFFGDIPIHAPVYLAVAL-----LHRTLNPEETSVVRKIG 351  
DB 282 WIER-----VVALDGMVSLHOPREHGPILLVMMLNPRHQHTVEBDDTRVQOGLS 335

QY 352 TAIQALNFQYLRLLQSLASGNDCTSTACCVYGLLSFVLTSEL-HTLQNDQDIIT 410  
DB 336 RAIKLGVEFELHAIYSHMFQKQSLTARIYKSIIRHNGFLQLQPDADSEIASHNINYL 395

QY 411 ACEVLAADSLPLFNGTR--PISGL-GIILDSVCCMPHLLSLPILQLBALVS-GKSTXK 466  
DB 336 LSELTSPITATEFKIDHPRALFNLCLE---PVEBFPKPAVALASASSQNK 451

QY 467 KYVSFLDKMSFYNELYKKEPHDI-----SHEDEGLAMRQRTKLLYPLGGQTNLRIPOGT 521  
DB 452 YIGLELLEMLPYTELITQNGQYERKKAASNDEDEIMCHD-----YTSSKINFTIPRT 506

QY 522 --VGVMLLDRAYLVMEYSYSWTLFTCEIEMLLHVSTADVI--QHCQV---KPII 573  
DB 507 KLIIVADLNHQRTY-VHPSGTVYFNAIHHEINLELDAQVSTLNSERVQRIATGKLYLA 565

QY 574 DLVHAYISDLSIACCLPITRIIMLQRLTVVSPPDVAVASCNCLTVLAANPAV 633  
DB 566 VAVRRIQOPHDITAAKVPH-TEMVPDVLKPRVAVNPVDVLVQCLANCTALVPLPQEI 624

QY 634 WTDLRHTGFLPFAHAPVSLISQMIABGMNAGYGNLNMNSHPOQSGYGVTTIAFLIYT 693  
DB 625 YTRIINMLIPSVNNANLTYQYANGVGDLSLIGYILMNVARNAGRYIILAAVYPLKT 684

QY 694 LVKQGLSTQSGD--VPCVMFYKEMLPYHKRYNSHGVQBOIGCLIELIHAIYLNCH 752  
DB 685 YTK--LGRDNVGVVELPGALFLLRVLFPYANNMRETSDBRYVVFVVQYIYEILQL-S 741

QY 753 EFDLHSSHPS-LQPLCTCSLAYTBAQGVINIMIGVDTIDMMAAQ-----PRSDA 805  
DB 742 EKLSDVARRYLRAQVSYSLNRNGWLLKIVGANGYLOAVERRSNMMLVPE---- 797

QY 806 BGGGQGLLIKTVKLAFSVYNNVIRLKPSNV--VSPLEQALSQGHANNT--IAVLAK 861  
DB 798 -----QQLAMTIQSMWTLIMQILKRIHINHYDLSPLSAIYTPQKQDTRILIPWVG 851

QY 862 YIYHKDPLPRLAIOQLKRLATVAPMSVYACIANDAAIRDAFLTRLQSKIDMRKYM 921  
DB 852 YMSNIFNPLPLTISCRLLRPAIFRPMGLACIDMEPDQIRHTPLERLADRIESDCLKA 911

QY 922 ILLEPLTAVETOPGLIEFLN-----LEVQDS---DGSK-----EPLSGMGC 962  
DB 912 VLEFVSSCVHKGPGLEAFKYNVKGKPDRLIADQGAASGSGSGRPRKTMINDIPY 971  
QY 963 LNAVLELIDQOQDQWYCPILHRAALAFHAIAMDQBRDSAMLVYRTKPEFMENTLSPLE 1022



Db 972 MBEFLAISNDPAK--LASPLLSR-IMSFLHLMKNNQALVKGLLQKNFMSSLCNPLF 1028  
 Qy 1023 GTLSPPSTSEBSILETCALMKITCLEIYYV 1054  
 Db 1029 S-----EOTS---SNVRSYSOLFNIIGIELFRV 1053

RESULT 11  
 ID 089083 PRELIMINARY: PRT: 190 AA.  
 AC 089083;  
 DT 01-NOV-1998 (TREMBlrel. 08, Created)  
 DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE Hypothetical protein (Fragment).  
 GN Name=BC025526;  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BALB/c; TISSUE=Spleen;  
 RX MEDLINE=99012997; PubMed=9798653;  
 RA Chu C.C., Paul W.E.;  
 RT "Expressed genes in interleukin-4 treated B cells identified by cDNA  
 RT representational difference analysis.";  
 RL Mol. Immunol. 35:487-502(1998).  
 DR EMBL: U89435; AAC36539.1; -  
 DR MGD: MGI:2446190; BC025526.  
 KW Hypothetical protein.  
 FT NON\_TER 1  
 FT NON\_TER 190  
 SQ SEQUENCE 190 AA; 21250 MW; EA00A3979BBD2B CRC64;

Query March 9.2%; Score 829; DB 2; Length 190;  
 Best Local Similarity 86.3%; Pred. No. 2.7e-47;  
 Matches 164; Conservative 5; Mismatches 21; Indels 0; Gaps 0;

Qy 1251 DQTHSLALGATEDKDSMETDDCSRRHRDQDQVGVGLHLAKELCEVEDGDSWLV 1310  
 Db 1 DQTHRLVSDSAEDKDMETDGCRRPHKQDQVGVGLHLAKELCEVEDGDSWLV 60

Qy 1311 TRRLPILFTLTLEVSILMKONLHFTBATLHLTLTARTQGAFAVAGAGITGICPL 1370  
 Db 61 TRRLPILFTLTLEVSILMKONLHFTBATLHLTLTARTQGAFAVAGAGITGICPL 120

Qy 1371 LSVYOLSTNGTAQPSARSKSIDASWPGVRLSMVLMEOILKTRNPLPALDPGVH 1430  
 Db 121 LSVYQLSSNGTQPTSTSRKSLISPSWGVRLSMVLMERLKTURNPLPALDPGVH 180

Qy 1431 QERTLOCINA 1440  
 Db 181 QNRTMQPFNA 190

RESULT 12  
 ID 09V6F6 PRELIMINARY: PRT: 1822 AA.  
 AC 09V6F6;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)  
 DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)  
 DE CG8771-PA.  
 GN ORFNames=CG8771;  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydriidae; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoeklin R.A., Galie R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.H., Blazer Y.H., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gaber G.L.,  
 RA Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier F.,  
 RA Burlis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz N.S., Gelbart W.M., Glasser K.,  
 RA Foster C., Gabrielian A.E., Garg N.S., Guan P., Harris M.,  
 RA Glodek A., Gong P., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heilmann T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kemison J.A., Ketchum K.A.,  
 RA Kimmel B.B., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M.,  
 RA Palazolo M., Peltzman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spler B., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svrtkac R., Tector C., Turner R., Venter A.H., Wang X.,  
 RA Wang Z.Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodger, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,  
 RA Yeh R.F., Zaveril J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers R.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195(2000).

[2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426065; PubMed=12537568;  
 RA Celniker S.E., Wheeler D.A., Krommiller B., Carlson J.W., Halpern A.,  
 RA Patel S., Adams M., Champe M., Dugan S.P., Frise B., Hodgson A.,  
 RA George R.A., Hoskins R.A., Lavery T., Muzny D.M., Nelson C.R.,  
 RA Pacle J.M., Park S., Pfeiffer B.D., Richards S., Sodergren B.J.,  
 RA Svrtkac R., Taber P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,  
 RA Weinstein G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;  
 RT "Finishing a whole-genome shotgun: release 3 of the Drosophila  
 RT melanogaster euchromatic genome sequence.";  
 RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).

[3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426070; PubMed=12537573;  
 RA Kaminker J.S., Bergman C.M., Krommiller B., Carlson J., Svrtkac R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.E., Rubin G.M.,  
 RA Ashburner M., Celniker S.E.;  
 RT "The transposable elements of the Drosophila melanogaster euchromatic  
 RT a genomic perspective.";  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).

[4]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426069; PubMed=12537572;  
 RA Miera S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
 RA Hradecky P., Huang Y., Kaminker J.S., Milburn G.H., Prochuk S.E.,  
 RA Smith C.D., Tupy J.L., Whitfield B.J., Bayraktaroglu L., Berman B.P.,  
 RA Bettencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,  
 RA Harris N.L., Richter J., Ruoso S., Schroeder A.J., Shu S.Q.,  
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
 RA Lewis S.E.;  
 RT "Annotation of the Drosophila melanogaster euchromatic genome: a  
 RT systematic review.";  
 RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).

[5]

RP SEQUENCE FROM N.A.  
 RG FLYBASE;  
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
 RN (6)  
 RP SEQUENCE FROM N.A.  
 RG FLYBASE;  
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.  
 DR EMBL, AF003821; AAF58471.3; -  
 DR Flybase; Fgsm0033766; CG8711.  
 SQ SEQUENCE 1822 AA, 208406 MW, EE2C45756B8AD70 CRC64;  
 Query Match 8.9%; Score 801.5; DB 2; Length 1822;  
 Beel Local Similarity 22.1%; Pred. No. 6.3e-44;  
 Matches 409; Conservative 319; Mismatches 769; Indels 351; Gaps 69;  
 QY 2 TRKSITVSLSPCRSSREIMTLIGRSALRELSQIAELKMKRRRLLEGSTYKPPSPS 61  
 DB 4 VEKSVIT-----DWKRLMPVSGIHETTPDUTVREELMNVASELQAGVLPQKPKNASS 56  
 QY 62 AE-----KYKANKDVASPLKEIGLRISEKLGDEBSQVLLQCYQEDRYGRDSVKTVL 116  
 DB 57 LELGLTLKEKKQEKLL-PFTE---RLQDLDESAGQWEILCYLTQETRGASASLTOLI 112  
 QY 117 QDEROSQALIKIADYYEERTCILRCVLLHTYFODERHPYRVEYACVDXLE-KEVLS 175  
 DB 113 STEYMAKIHEDIIRHYSLERWVVLKIVQLIVFHQVPMHPYRREYRAVEKITTIRLSD 172  
 QY 176 KYRQOFEEELYKTEAPTEWETHGNLM--TERQSRMFWQCIREQSMLEITFLVYAYEEMAP 233  
 DB 173 SYLDLSEILCEVPPKAKMAEGCFHSAEHLVA-WSEKNAKEITNEVHLILVLAELPMGL 231  
 QY 234 SDLLVLTWKFEGQSGRQTNRLNVDETW---DPFVDRIGYPSALILVEGMDIESLHKCA 290  
 DB 232 EQIKRIIPAACKQHSFGKQKS---YLDSDQPYHQEIRLSYSGLMVLCCLDFE----- 282  
 QY 231 LDDRELEHFOADGLICQMDCLMTFGIIPHHAPYLALMAL--LHNTLNPEETSSV- 346  
 DB 283 ---KEEKSDILIEKI-EDLQVDIASMYHRPHEGPLLAMMLRLRGITDADASSLKLC 338  
 QY 347 RKIGSTAIQVNFYVLTLLQSLASGNDCTTACMCGYGLSFLVLSLEHL-TLGNQO 405  
 DB 339 ROLGKRAVDLKCFTVQLHIAHSMWADDSMLRIYRTITVNOGYICDLFPDGGSCARYE 398  
 QY 406 DIIDTACVADPSLELFWGTEPTSGLGIIIDSVCGMPHLLSPILQLLRALVSGKSTA 465  
 DB 399 GIYELLCELVSWPHLAKOFCRSRELT-----HLSKALSTKA-----GGG 439  
 QY 466 KKVYFPLDMGF-----YNELYKAKPHDIVSHEDGTLMRQTPKLLYPLGGQTNLRIPOST 521  
 DB 440 NYVKSQALAPLILALRYDE-SQHKUREVDTEFEELASVQPO-----QIDFTIPAGT 491  
 QY 522 VGVQVMDRAVLYVREYSSSWTLFTCEIEMLL-----HYVSTADVIQHQRKPIIIDLVH 577  
 DB 492 SCTAIQHSQSCPMFRFPVNYFDALHHEINCLRETGHLDGPSESRIRNVEAGRLFE 551  
 QY 578 KVISIDLSI---ADCLLPITSRIYMLQRLTTVISPVPDVASVCNCLTVLAANPAKV 633  
 DB 552 SAVKLSQSIGISAMVHP-TEMCDVLHTFKSVQYPPVGLSSCLINVTALLPLVDEBI 610  
 QY 634 WTDLRHTGFLPV---AHVVSLSQMSISAEKNAGGYGNLNMNSQPOGEYVTV---IAF 687  
 DB 611 FSRISNHLILFPVSGSHYDFGMVYANANGVFESRFLGSVIDNVKKEERYEFLLSYIGF 670  
 QY 688 LRLITTLVKGQAGSPOSQGLVPCWVFKEMLSYHKMKRYNSHGVREOIGCLILELHAI 747  
 DB 671 LRAYSNLKRN---QIQWEILPGLIFLKDVPFHLHTHFSQVSEKNTIYEILSPICDI 726  
 QY 748 LNL-----CHETDLSSHSTPSLOFLCICSLAYTEAGQVYINIMIGIVDTIDVMAAQ 799  
 DB 727 LDFLFTAKESNCKQREL-----LVKVCVYSILMLLENGLILLRVGVGNNAVQYTMLE 779  
 QY 800 PRSDGAEQGGQGLIKIVKLAFSVTNNVIRKRP---SNVVSLEQALSGHAGNNL 855

DB 780 ---TNMQOQPHGIML---VRLSMRIIMQLRLKEEYVGNSETLSLEALITYQPKORDYL 835  
 QY 856 ---IAVLAKIYHKHPALPRALIOILKRLATAPRPSVVA-CIGNDAAIADAPLTRQSKI 913  
 DB 836 RILPTVCSMSNIFPRMPLPILSCRLKRILOFNNSLACIMEDQILHTTMOQKLPDEL 895  
 QY 914 EDMRIKMILEFLTVAVETOPGLIELPLN---LEVKQSDSKER--SIGMMSCLHAYL 967  
 DB 896 BSDSIKIMILELVADACIAKQPVTEAPFKVNVALDKRGRSPSKOCVPRIGR-SITYVNR 954  
 QY 968 ELIDSGQODRWYCPPLHRAIALFALMODRDSAMLYATKPYKMENTSPPLFOTLSP 1027  
 DB 955 DFLDLQVDPPLTIOALPAKIMTIFHSW--KHNLQMLY----- 991  
 QY 1028 PSETSPSILETCALIMKICIEIYYVVGKSLDOSLQTLKKSIEKRPAYSGYVKSIA 1087  
 DB 992 -----DILLNITISIEV--TGNGNNAALLDVNKFEPQKFGPMLNTYFNP 1037  
 QY 1088 VHAETEGSSCTSLLEYQMLVASAMRLIITVHADIMHTDVSVRQLFLDLDOTKAL 1147  
 DB 1038 KVPAYKONLSSBDHPDWICCLQAPKDLVI-----LKKQPKF 1075  
 QY 1148 LVPASVNCCLRGSKCTLLLI-----LIRQWR---ELGSDV 1182  
 DB 1076 VTIPESQ--FKLMQKCLVIVDRSNYLEDKRPILIAELVYFILPFGHAYTDSLBEQO 1133  
 QY 1183 ELIGPLTEILSGVLA--DOOLMEK-----TKAKVSAFTVYQMKEMKSDIPOYSQ 1233  
 DB 1134 TLMDDLQLMNRICACTEDQVRAKACLAIVTK--THLYTDLILDSGIA-LRPLNS 1189  
 QY 1234 LVLANVCEILOEBVIALFDQTHSIALGSAEDKDSMETDSCRSRRHQRDQVCVYGLTL 1293  
 DB 1190 VVGICSELEQW-----ENSVLSKSGQANNDSDSKSTNS-----LILCML 1234  
 QY 1294 AKELCEV-DEQG-DSMLQVTRRLPLPTLLTLEBSLBMKQNL-----HPEBATL- 1341  
 DB 1235 LKAVAITIFHNDGPGW-----DLRP-----VSVLFQRLVACVSRTPLPFLSKOYLS 1280  
 QY 1342 -HLTLTLARTQOGATV--AGAGITQSICTPLLSVYQ--STNGAQTGPSASRSKSLDAPS 1396  
 DB 1281 VQLDVLIVPAKHCVSFELCDVGRYIMKLLPRLRLQSKHETKTTAADAEGWYEQ 1340  
 QY 1397 WGVYRLSMSLMDOLKTLRYNPLPEALDPYGVHQERTLOCLINAVTYQST--ACLEBAD 1454  
 DB 1341 WMPVYARGIELVITIIYBKCKCPLEDAFOGIAHVPLEDAL--LLSKOSLEPSAMYLTK 1398  
 QY 1455 HTVGPILDSNPKEMW---HFHLPOLMRDIOVNLGYICQACTSLHSRKLQHTLQKN 1510  
 DB 1399 AAVNLVASLTERHHEMKQSDSLANIMRAVO---SLLCHT-SSLFHQOKMLKILL--- 1450  
 QY 1511 GDGLPSAVAQVORPPSAASAPSSKOPADTEA-SEQQALHTVQYGLIKLSKTLAL 1569  
 DB 1451 -----AGRSQLEILASTRLIYDBELISACNDLTDIIISCVRL 1490  
 QY 1570 RHFTEDVQIILDOSLDAEYNPLFALSFTPTPDSR-VAPSGTLATVNV--ALNN- 1624  
 DB 1491 LRPSDLMEILCCSAYESKSIILDDYFGAPKLBENLTLTFGLVLMVNLVYKALNNQ 1550  
 QY 1635 -----LGEIDKCK-----EPLQAVGLSTQABSTRILKSL 1655  
 DB 1551 NHGFSVPPLNSLPVNEHSGDDDBEVCVGNQTNRTFSPRLSN-VSIST--GTCPASBIL 1606  
 QY 1656 -----MPTMENCQYLLISQARVYLDPAVHPDRKQRMKQBSSEL 1695  
 DB 1607 SNMGQCLLALBEHLMLVASOALCIIRSPULETLMKQIVRDISNEL 1654  
 RESULT 13  
 OEDRN9 PRELIMINARY; PRT; 1997 AA.  
 AC OEDRN9, 01-OCT-2004 (TREMELrel. 28, Created)  
 DT 01-OCT-2004 (TREMELrel. 28, Last sequence update)

DT 01-OCT-2004 (TREMBLE rel. 28, last annotation update)  
 DE C7orf14-like.  
 OS Brachydanio rerio (zebrafish) (Danio rerio).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
 CC Cyprinidae; Danio.  
 NC NCB1 TaxID=7955;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Ameternam A., Nissen R.M., Sun Z., Swindell E.C., Farrington S.,  
 RA Hopkins N.;  
 RT "315 Genes Essential for Early Zebrafish Development";  
 RL Proc. Natl. Acad. Sci. U.S.A. 0:0-0(2004).  
 DR EMBL; AY648720; AAT6038.1; -;  
 SQ SEQUENCE 1997 AA; 223244 MW; 933087C347CFC351 CRC64;  
 Query Match 4.1%; Score 372.5; DB 2; Length 1997;  
 Best Local Similarity 19.1%; Pred. No. 3.9e-15;  
 Matches 420; Conservative 296; Mismatches 799; Indels 685; Gaps 98;  
 QY 1 MIRSKITSVLSFCSSRELWTLIG--RSALRELSQTEALNCHMRRLGSLYKRP 57  
 DB 1 MATQAVNVSAGSLWQPLKELWEAVEGAIWRQPESEVHLIDQLKGRKPYF--LNLKXP 57  
 QY 58 SPSSAEKVKANDVAPLKEGLRISKFLGLDEOSVOL-----OCYLOEDY--RGR 109  
 DB 58 PKSAEGRKVRK--AS-----TEGAIQGGGAKLLPQOLISEAFILSDLPDGL 106  
 QY 110 DSVKTVLQDERQ-----SQAIIKIDAYYEERTCI--LRCVL----- 145  
 DB 107 AALBELLAGEHQHPFGLTRGLVAVL--LYWDGKRCVANSLSRLSIOSRHKFTLIDSLP 164  
 QY 146 ---HLITTFQDBRHRVREYADCVDKLELVS--KTRQFEEIKYKTEPMETHTGNLMT 200  
 DB 165 ELVNLTFRTBEL-----MAQGLTKQILNLSVSVTEFERLQKERG--LGNEXH 213  
 QY 201 EROVRWFOCUREOSMLEIFLYAYFEMAPSDLVLTQMKFKEGFGSRQTNRLVDE 260  
 DB 214 RKEVSDLLKEC--ROSLAECLPANAQSPPLGKDTLALI-----GHLEMYTA 258  
 QY 261 TMDPFDRIQVFSALILVEGMDIESLHKCALDREELHQ---FAODGLICODMCLMLT 316  
 DB 259 EADGSLDSVNLAVALVSLDVSFLDQ--GTEDREDLQALPLTEKQVVAVHSRLVSG 317  
 QY 317 FG-DIEH-HAPVLLAMALIRHTLN--PE-----ETSSVVRKGGTALQIANVQYLTR-LL 366  
 DB 318 QGWKLPGLQAVVQVLMALSTRALSQLPQCALVEFEADEALADQLAGVFLPFLTEGVL 377  
 QY 367 QSLASGMDCTSTACMCVYGLLSFVLTSLHTLIGNOD-----IITDACEVLADPSLPE 422  
 DB 378 GSDGFSQEEFYRLRLHSLITDPLA--LMEKVKQLNRADEDARLIHMLQW----- 427  
 QY 423 LFMGTEPTSGIILD-----SVCGMF----- 444  
 DB 428 ---GSEPPSSLRKQIDHMLILGEFYSKDPFLELALFEWCPSELSQHTSLGSLGVFL 484  
 QY 445 ---PH-----LISPLLQILRALVSGSKTAKKYVSFLDKMSFYNELY 482  
 DB 485 ORPFTTEVVLKSFVROMGDLPLATLYIPLYLRMKKGLANPQCSHVCFSILKT----- 536  
 QY 483 KHKPFDVSHGELTMRKOTPKLLVPLGGQTMRLRPGQTVGQVMDDRKXIVYRWESYS 542  
 DB 537 NGAPF-----GGRNQ--AGVSGS-----LVSWEHFFFS 562  
 QY 543 WTLFTCEIEMLLHVVSTADVIQHCORVKPIIDLVHKVISTDLSIADCLPISRIYMLQ 602  
 DB 563 LMLYH---ENLRADVPASDSTQY--RHLPIRGIQTQEL-----EGLTAFQ 603  
 QY 603 RLTVVIS-----PYVDVASVCNC-----LTVLAA--RNP--AK 632  
 DB 604 LITTIITSENARLALCEHPQWTQPVVVMGLQCSQVPLKAQVHLVLAFAKSPRIAS 663  
 QY 633 VMTDLRHGFLPFAVHPVSSLSQKISABGMNAGYGNLNMNSEQPGEGEVITAPRLIT 692

DB 664 LMQSLLEYQIILQTVYIPIQORQAGIEV-----LNEIESSCEEPPLFRACHILS 713  
 QY 693 TLVKGQL-----GSTQSGVLPCVWFVLEKMLPSYHKVRKYNHSGVBOIGCLILEIHAI 747  
 DB 714 TLVESALPVTIGAGIRAPGFQYIDFLDSVFLAPPTAIRRSARKMEVAEVLVEFRL 773  
 QY 748 LN-----LCHETDLSHSTP-----SLQF-----LCICSLAYTEAQTVIN 783  
 DB 774 LREYEPQSDPLPEVNEVIGSGVPAHMRPGHSLMHLNDSPTSLC-----LN 822  
 QY 784 IMGIGVDITDMVMAAPRSDGAEQGGQQLIKTVKLAFSVT--NNVIRKPPSNVSPLE 842  
 DB 823 LLEBGAQOLD-TYAPPEKKQLESAYLHCLCLELPLQKQVTFMDLRESQTSILVSPLE 881  
 QY 843 QALSQHGAGNNL--IAYLAKIYH-KIDPALPRLAIOQLKRLAVAPMSV----- 890  
 DB 882 QLLQGVSAQSRKADHTITARIYLYHSSNPDAFQSAKILRIRITRYPNIOARLVGDFTHD 941  
 QY 891 -----YACLGNDAA--AIRDAFLTRLQSKIEDMRKVMILEFTVAVE--TOP 934  
 DB 942 QAVSERLMAQFVECLDSEBAQGVTTNDSSEKRAVIRH-ETKHILMLITSLEITCP 1000  
 QY 935 GLIELFLNLEYKDGSDGSKFESLGMM--SCLAAVLELIDSQQDRYCPPLHRA-- 987  
 DB 1001 NLGLVILGYEVKPVSTNLQDPVILGCPSCIAHILSLQGRSDSRSG-PGLIKQAPQL 1059  
 QY 988 ---AIFPALWDRDSDAMLVARTPKFRENLTSLPFGTSLP--PSETSPSI--LET 1039  
 DB 1060 AELCYQVYIYQACACDSTGSPTRRYLRTSODF-----LFSHLQHLPVLSSENEALASQ 1112  
 QY 1040 CALIMKIICLEIYVYVVGSLDQSLDTLKKSIEKPAWMSGYVSLAHVAFETGSSCT 1099  
 DB 1113 MSWLMKTAIRKATV--SLNRQSHYQRLNL-----LDDQPHILHTQGE--T 1157  
 QY 1100 SLELEYQMLVSARMLLIATTHADIMLTDVVRQ-----LELDVDTKALLLVPAV 1154  
 DB 1158 GMBEENRVSQFLQGVTSKYVRRLSVLDAIDFQABELLQDLPFERTQIEQYI--T 1214  
 QY 1155 NCLR--GSMKCTLLILRLQMKELSGVDEI--LSPLELLEBVLQADQDMKRTAK 1209  
 DB 1215 NCEHNEQGHVTCNKL--LHRVLVAEINALQGMALIGRPMLMEVNSVLAQVVERNR 1273  
 QY 1210 VF-----SAPITVLMKEMKVS DIPQVSQLVNAVCEYQOEVI----- 1247  
 DB 1274 RSLAKRRALDSKWSLVEITLILACPSSELPADQROLITRDLLDHDKVLSDAAGELMP 1333  
 QY 1248 ---ALPDQTH----- 1255  
 DB 1334 IVAGAVFTLTHALSGSVSEQGVGPBEGSSGSPASINSALHLIRKLDLPILCTGGGF 1393  
 QY 1256 -----SLAGSATEDKDSMET-----DDCSRRHRDQGVCTGL 1291  
 DB 1394 QELRAHLYGALLYLQIAQKEEPEPTLQTSMSMERLTAPEBCFGLKQRENIISIESYCT 1453  
 QY 1232 HIAKELCEVDEG-----DSWLQVYRR-----LPIPLFLTLTXS 1327  
 DB 1454 ALMEVYAGDADCGHEIGMLALAVLDVSLSDRQOMLVILCNSGYLRVLVSLKQDDVS 1513  
 QY 1328 LR-----MKONLHFEATLHLTLTLARTQOGATAVAGAGITOSI--CLEPLSVYQIS 1377  
 DB 1514 LQTLTTPQPLPKPIYTESKALTLRYAKTRQGMELLRCGLVAGQVBC--QVFHML 1569  
 QY 1378 TNGTAQTPSASRKSIDAPSWPGVYRLSMSLWEQLIKTRYNFLPALDFGVGHQERTLO- 1436  
 DB 1570 PONDALRVFGGD-----PS-----GRIPSPPL-----RRYQOI 1597  
 QY 1437 CLNAVRYQSLACLEADHTYGFILQLSNPKGW--HEHILPO-LMRDIOVNLGYLQCY-A 1491  
 DB 1598 LIPTRLMQVILITSTTAHQQAQVLAQVL-----QWLIHSDVLAQSLIHGDQNMGSLQDELS 1652  
 QY 1492 CTSILHSRMLQHYLVONKNGDGLPSAVAO-----RVQRPSPSA----- 1529

DB 1653 LITAIISKALPGALB--MGQEIINSAIMELQGIHGFQROSLSLVRLVGTDRARYLKQ 1710  
 QY 1530 ---SAPSSSKOPADTEASEQOALHTVQGLKILISKITALAHFPPDVCOILLDQSLD 1586  
 DB 1711 IEDTVSPSNTAEKREDEVMQOCANIM-----EYCOITLLQSSS 1751  
 QY 1587 LAEYN--LPAISFTTP---TFDESVAPFG--TLATVNAVNLMLGELDKKEPLTQ--- 1637  
 DB 1752 EAGPSICLPSPASBPADVSIIPSAKVSLGLVILLNLSGSDPFRYHDSHQSLNKLERY 1811  
 QY 1638 -----AVGLSTQAEGRITLKS-----LIMFTMENC 1662  
 DB 1812 EQLPEELKELCOGLVSGSGVEKISSVQRRVLAKEHLVOLVNNRAKMLALCSDIITETC 1871  
 QY 1663 FYLLIQAMRYL-----RDP--AVHPRDKQRMQOELSELSTLSSRYFRGAPSSPA 1715  
 DB 1872 LFIYMHLEFYLYCYPTTDKDSLIPGYROQLSLKLDYFQMLNDVSY-----Q 1922  
 QY 1716 TGVLPSPQO--KSTSLKASPSQEPFL--IQLVQAFVRHMOR 1753  
 DB 1923 IALCAPAGSRGLSVSRVSGQDLQGLQSDMSSSPSELPQR 1962

RESULT 14  
 Q81M71 PRELIMINARY; PRT; 2025 AA.  
 ID Q81M71; 081M71;  
 AC 01-OCT-2002 (T-EMBLrel. 22, Created)  
 DT 01-OCT-2002 (T-EMBLrel. 22, Last sequence update)  
 DT 01-MAR-2003 (T-EMBLrel. 23, Last annotation update)  
 DE Hypothetical protein OSJNA0034B05.4.  
 GN Name=OSJNA0034B05.4;  
 OS Oryza sativa (japonica cultivar-group).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzaceae; Oryza.  
 OC NCBI\_Taxid=39947;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Wang R.A., Yu Y., Yang T.J., Nah G., Soderlund C., Chen M., Kim H.-R.,  
 RA Rambo T., Sasaki C., Henry D., Oates R., Simmons J.,  
 RA Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AC12145; AAM46049.1; -  
 DR Gramene; O81M71; -  
 KW Hypothetical protein.  
 SQ SEQUENCE 2025 AA; 224968 MW; 9F69BD4AA1FE89A0 CRC64;

Query Match 4.1%; Score 368.5; DB 2; Length 2025;  
 Best Local Similarity 18.1%; Pred. No. 7.3e-15;  
 Matches 384; Conservative 327; Mismatches 652; Indels 755; Gaps 98;

QY 25 LGHSALRELSQIEA---ELANKHMRLLLEGISLYKPPSPSAEKVKANKVNAS----- 73  
 DB 45 LDRAAASBPVPALMERIKAHAMLRGSVMFVKPDSARGLADSEVVVGGHRLAVKP 104  
 QY 74 ELKELGIRISKFTGLDEBSVOLLOCYLQEDYGTBSVKTVALODEQSQALILIKLADY 133  
 DB 105 ELKAAALRISKCNMLDEVQSYILV-----KTSINTPTALVAD--TESEFLAVSVQY 154  
 QY 134 YEERTCILRCVAILLTFQDERHPYRVEVADCVDKLEKELVSKYQOFEF-----LYKTE 188  
 DB 155 YLERQCLIKCIRIRIIFYHND-----CSDSIDAVAEASVLYREVEGRLLSIVADSL 206  
 QY 189 APTWETHGNLMTERRQVRWFVQCLREOSMLEIIFLYVAFEMAPSD--LIVITKPKFE 245  
 DB 207 ASAPSVYKG--AEILTS--WLEBTLIRINLIFDLIFLF--YDNLSCRNGGLMIMLCISIFKO 262  
 QY 246 QGSGSQTNRHLYD--ETMDPVDRIQIPSA---LIVEGMDISELHKCALD---RREKH 298  
 DB 263 MUGSYDVGVFAVSVKNSF---HYAKAQLPFIILQTLDFESLIMVDEVPFSGGYS 318  
 QY 299 OFAODGLICQDMQIMLTFGDIPHHAPVLAMALIRHTLNPBEETSVVRKIGGTALQNV 358

DB 319 TFSVVDILEMDEVSKLPEFAVESGPLILAMAVF-----LCLVMSLPSSNTNLP 369  
 QY 359 FOYLTRLQSLASGNDCTTSTACMCVYGLSLFVLTSLBLHTLGNODDIIDTACEVLADP 418  
 DB 370 VSGFRILKTFIS-----AVVASY----- 388  
 QY 419 SLPFLMGTSEPTSGLOIIDSVCMPHLLSPLLQD--LRAVSGSTAKVYSFLDKMS 476  
 DB 389 ---EISQYEDDS--LGMILNIIICEVYDGEBSLQCMQWDSKDFIDGPIR-----FVYLERNN 440  
 QY 477 FVNELY--KHKPRDIVSHEDGTLNRQTEKLYPLQGGQTKLRIPQSTVGQVM--LDDRAY 532  
 DB 441 GTVTLAVPRSDTDVNNYHQI--EHSPISTIFIGGT--TPGSHGILVLEBDVA 495  
 QY 533 LVREYSEYSSWTLFTCEIEMLHVSTADVIQHCQKVPFIIDVHKVISTDLIADCLP 592  
 DB 496 LVNME-----DLCLALHAKSLAVQASQ---LGIYDHNVRIDIAKIFC--- 537  
 QY 593 ITSRIYMLQRLTT--VISPPVDVIAVCVNCITVLAARNPAKYWTLRHTGPLEPVAAHV 650  
 DB 538 --TSIFKYVEDFNNAFCVMSKTLGMLAEMLSCV-----PYHFNVALDCGF--FIRQ-- 584  
 QY 651 SLSQMIISABGANNAGGYGNLMMSEQPOGEYGTIAFLKILITLVKQL-----GSRQSQ 705  
 DB 585 ---SGVASSDWMISGLAARMLPATSESDSCS-----SLTTVLDPALQVLRKGAAD 635  
 QY 706 GLVPCVMPVLKEMLPYHKRYNSH-----GVREQIGCLILEL-- 743  
 DB 636 IISFIIIFSVQIYIMVNNHMKIKYSRMKLTILKVPVLKSCIOVKSFPSSLDGIIWEILL 695  
 QY 744 ---HAIL--NICHTEDL---HSSHTPSLOFLCISLAYTBAGQTVINIMIGIYVDTID 793  
 DB 696 YDSHSVLAHLILMSQTQLSHSGSYCHDKOI-----BDIQVLCCGPFIVF 744  
 QY 794 MVAAPAPREDGAGGQGGQGLIKTV--KAPSTNNVILKPSNVVSPLEQLSQRGAH 851  
 DB 745 YMLSNLP-----EREKSKRMRLRVGSKSAFRYGVFVAVKTLA-----FRH 786  
 QY 852 GNNLIVAVLAKY-----IYHKDPALEPLAQLKRLATVAPMSVYACIAND--- 897  
 DB 787 SCNIIIVLSBFSQPSIEMKTCIIFH-----LOLAPIKILS---QSHYAKSNGSDNN 834  
 QY 898 -----AA-----IRDAFLTRLOS- 911  
 DB 835 RTSNKEPMDNDNDIFLKAIOVAARVPSMLCFYAYKAQOLMENAYFVVGSGEIMRLQTS 894  
 QY 912 -----KIEDMKIKVMIIEFLVAVETOP-----GLIEFLNIEVDGS----- 949  
 DB 895 ISCLIDEVQKVEVVAIFMLISAAARYQARADVALVWLRVHLPGQSTRAQADSDNSAH 954  
 QY 950 -DGSKEPSLGMSCLHAVLELI-----DSQODRYWCPLRLHRAAIAFLHALMODRDS 1002  
 DB 955 BOSSKTFVLPSSGNRLVEQIIGYGRSTELMDR--SPSILSGVLDLKLALMESGAQF 1011  
 QY 1003 AMLV--LRTKPKEMNLTSPLEFTL--SPPSST--SEPSILBTCL--IMKIIIEIY 1053  
 DB 1012 IYILEKLRSSRTFEMNLSCCIRAPASYPIDSVETVDKSKSLKCYCGITFLIMSYELFL 1071  
 QY 1054 VVK-----GSLDQ-----SLKDTLKRSIEKRPAYMS 1080  
 DB 1072 QGKILITETKTSIDPAVPGSKQKPSVAPCPSDIVLWKPFSTIEMBDIVNHLSS-----N 1124  
 QY 1081 GYKSLAVVAET-----EGSSCTSLIEYQMLVSA-----WRML 1115  
 DB 1125 GYQNDI--LHRAKVASGLCTIIRLLTKLSSGDTGSLSVLKVQIQLISSKLTMCAGATTL 1183  
 QY 1116 I-----IATY-----HADIMHLTDSVVRQQLFLDVLDGTAKALLVAVSVCLELGS 1161  
 DB 1184 CGSPRIASISGNOVHWHRRHGRDAOSTDA-----NAPFGIGVVTTP-----CI----- 1227  
 QY 1162 MKCTLLIILLRQKRELSGVDE-----IIGPTEIL----- 1192  
 DB 1228 --CSSTVTICPSWILQWILFMDENPLHDIWLKTLHIALEGRGPIFGPQBLICFLAF 1285

QY 1193 -----EGVQAD-----QQLMEK----- 1205  
 DB 1286 KVFENHPSBQLOKSPPAANGVSLFDVPHIRDELGLMNHNSDMKTYKVAERQDLIMKA 1345  
 QY 1206 -----TKAVSFATVLOM-----KEMKVSIDP-----QVSQVLANVCETLOEVE 1246  
 DB 1346 NLMKQVADKCALMSFTIFLVCVGTSGTSYKFGFBGGISITTTQSAVRACKSKISQSV 1405  
 QY 1247 IALFDQTHS-----LALGATEDSDS-----METDDCSRS 1277  
 DB 1406 DSLPEVNSGVLPPLSGQVELLTLTRILLDHAKQSKSRHLYPVITLMTSGASYS 1465  
 QY 1278 -----RHQRDQ-----VCVGLHAKEL 1297  
 DB 1466 FLFNMPSSPALQPVKSLVLLSLFEPTIKKVDKDSQEDVNI FGLISLMSLTPVL 1525  
 QY 1288 CEVEDGD-----SMLOVTRRLPILPTLTLTLEYSLRMKN--LHFEATL 1341  
 DB 1526 CKLASREYFDLALMSMDIILKGFPSNVWVILQGHF--RLQVILQKQSGALLCTQVTL 1584  
 QY 1342 HLLTLTAQTQAGATAVAGAGITQSIQCLPLLSYVQSTNGTAQTSPASRSLDAPSWPGY 1401  
 DB 1585 NFLTMTGRKQKAKILQSANIFAFIKV--LLS--QNSLDSCLRNSLSTQTKQVKIW---- 1637  
 QY 1402 RLMSLMEOQLKTL-----RYNLPBALP-----VGVHQRITLQCLNA 1440  
 DB 1638 GLGLAIVSLKNCMDDISRNSVANSSTISFLSGQVPLMSSYLSAOSVNHQSKRTLOK 1697  
 QY 1441 VRTVQSLACLEADHVTGFILOLSNPKMHFPLPOLMBDIOVNLGYLOACVSLHSRK 1500  
 DB 1698 SQT--SLSLSTLEN--LTLTLCTLAKHPRDTGKGVSELS-----RE 1738  
 QY 1501 MLOHYL-----QNKXGD-----GLPSAVAGRV--QRP-----SAAS 1530  
 DB 1739 ITHLAFISRGSESRGDSPPNNLSFGCPPIKEBKMLBEPPLISKYGMFPAPASCTL 1798  
 QY 1531 AAPSSSKOP-----AADTASRQ-----ALHTVOGLKIL--SKTLAL 1569  
 DB 1799 STPSVSGPPNAGLSLIVRDKNPDSMKQRTPTMLAVQIYIATLIMKFLCSQKEXAV 1858  
 QY 1570 R-----HF-----TPDVQILLDSLDLAEVNFPLSFTPTPTDSVAPSGFT-- 1613  
 DB 1859 RRAELFLDLAHPPELPPMDILHGLQD-----VSVIVTELVANVSTALNTB 1907  
 QY 1614 -----LLATVVALNM 1624  
 DB 1908 TERVQCLLVILETSLYM 1925  
 RESULT 15  
 N205\_HUMAN STANDARD; PRT; 2012 AA.  
 ID N205\_HUMAN  
 AC Q92621, Q86YCI, 43, Created  
 DT 29-MAR-2004 (Rel. 43, Last sequence update)  
 DT 05-MAR-2004 (Rel. 43, Last sequence update)  
 DT 05-JUL-2004 (Rel. 44, Last annotation update)  
 DE Nuclear pore complex protein Nup205 (Nucleoporin Nup205) (205 kDa nucleoporin)  
 GN Name=NUP205; Synonyms=C7orf14, KIAA0225;  
 OS Homo sapiens (Human)  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OC NCBI\_TaxID=9606;  
 OX [1]  
 RN SEQUENCE FROM N.A.  
 RP TISSUE=Bone marrow;  
 RX MEDLINE=97191544; PubMed=9039502;  
 RA Nagase T., Seki N., Ishikawa K.-I., Ohira M., Kawarabayashi Y., Ohara O., Tanaka A., Korani H., Miyajima N., Momura N.;  
 RT "Prediction of the coding sequences of unidentified human genes. VI. The coding sequences of 80 new genes (KIAA0201-KIAA0280) deduced by analysis of cDNA clones from cell line KG-1 and brain";

RL DNA Ref. 3:321-329 (1996).  
 RN [2]  
 RP SEQUENCE OF 1181-2012 FROM N.A.  
 RC TISSUE=Skin;  
 RX MEDLINE=2238257; PubMed=12477932; DOI=10.1073/pnae.242603899;  
 RA Strausberg R.L., Feingold B.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loguettano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
 RA Bosak S.A., McGowan P.J., McKernan K.J., Malek J.A., Gnaratine P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,  
 RA Villalón D.K., Wuzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahy J., Helton B., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko A., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butlerfield Y.S.N., Krzywicki M.T., Skalska U., Smalins D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
 RN [3]  
 RP SEQUENCE OF 85-92; 572-581; 1452-1461 AND 1494-1502, AND  
 RP CHARACTERIZATION.  
 RX PubMed=9348540;  
 RA Grandi P., Dang T., Pane N., Shevchenko A., Mann M., Forbes D.,  
 RA Hurt E.;  
 RT "Nup93, a vertebrate homologue of yeast Nup96, forms a complex with a  
 RT novel 205-kDa protein and is required for correct nuclear pore  
 RT assembly.";  
 RL Mol. Biol. Cell 8:2017-2038 (1997).  
 CC -1- SUBUNIT: Interacts with NUP93.  
 CC -1- SUBCELLULAR LOCATION: Nuclear pore complex (Potential).  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation.  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; D86978; BAI13214.1; ALT INIT.  
 DR EMBL; BC044255; AAH44255.1; -  
 DR Genew; HGNC:18658; NUP205.  
 KW Direct protein sequencing; Nuclear protein; Transport.  
 SQ SEQUENCE 2012 AA; 227918 MW; FF92PCB40B723F16 CRC64;  
 Query Match 3.2%; Score 292.5; DB 1; Length 2012;  
 Best Local Similarity 18.9%; Pred. No. 8.9e-10;  
 Matches 416; Conservative 303; Mismatches 743; Indels 741; Gaps 104;  
 QY 41 NKRMRLLLEG-----LSYK--PPSSSAEKY-KANKDVASPLKEIGLR- 81  
 DB 26 NLMTRQPAVAVHLDKLKHKPFIPLKPNPKVQVQHEKQVAKSTEGVALQGGQGTBL 85  
 QY 82 -----ISKPLGLDEBSVQVLOQC--YLQSHYRG--TSDSYKTVQIDERSQALL 127  
 DB 86 LPEQLIKENFIISDFDIEGLAVALLELAGEHQPHFPLTGLVAVALL- 134  
 QY 128 KIADYYEERTICLVHL-----TYFODERFPRVRYADCV 167  
 DB 135 -----YWDKRCIANSIKALIOSRKQTTLSLSPASMTTRFDLMEQGLTYVAVL- 188  
 QY 168 KLEKEIVSK--YRQFEELYTEAPTWETHGNLMTEROVSRWFOCLREOSMLLEITLY 225  
 DB 189 -----LVSGIDVNNRERKQIRERGLSGSKG-----RKEVSDIKEC--RQSLASBLFAW 235  
 QY 226 YVFFBAPSDLLVLTVMFGQGFGRQTRHLVDEIMDF--FVDRIGYSALIVGMDI 283

Db 236 ACOSPLGKEDTLLI-----GHLERVTAEANGSLDAVNLALMLALCYCFDI 281  
 Qy 284 -----BELHICALDRE-----LHOPADGLICODMCMCLTEBDIDH-HAPVL 327  
 Db 282 SFLGSTEREDMDIHOPLTEREQYATATHSRLQDSQLMK-----LGLQATVA 330  
 Qy 328 LAMALLRHTLN--PE-----ETSSVVRKIGGTAIOLNVOYLTRLLOSILASGNDCTST 380  
 Db 331 LAMALLRGTISQPLPDVLTALAEFTADEAMALADNVFLP---LMSVAVSEFYQEEF 387  
 Qy 381 ACMVCYGLLS--FVLNLSBLHTLGNOD-----IIDTACEVLADPSLP----- 421  
 Db 388 YLRVRNLTITDPLALPMKVKQLNRADBDARMIHSMQGNBPISLRDLHMLLIG 447  
 Qy 422 -----ELFMGTERT-----SGG-----IILDSYGMF-----P 445  
 Db 448 ELYKNPFLHLEALFEMWCPTEPLQPTLIMGSYLGVANRPPORQVYLSKFRQMGDLBP 507  
 Qy 446 HLLSPLOLRLALVSGSTAKKYSPFLDKMSFYNELYKHKPHDVISHEDOTLMRQTPKL 505  
 Db 508 TYIPLKRLQGLANPQCHNYCFSL-----534  
 Qy 506 LYPPLGQTNLRIPQGTGVQVMDLDRAYLYRMEXYSYSWTLFTCEIEMLHVSTADVIH 565  
 Db 535 --KVNQSHVENIQAAGSP-----VSMHFPHSLMXX--EHLKDLPSADSVQY 581  
 Qy 566 CQWVRPDIILVHKVISTDLSIADCLPIRSRIYMLQ--RLTVIAP--PUDVIASCYN 620  
 Db 582 --RHLPFRGITOK--BODGLIA--FLQITSTIITWSENARLALCEHPQMTVVVILGLQ 635  
 Qy 621 C-----LTVLAA--RNP--AKWTDLHGTFLPFVAPVBSLSOMISAEAGNAG 665  
 Db 636 CQIPPLKABLKTLLAFKSPBIASIMOSLEYTQLOTVRIPSRQALGIVE----- 690  
 Qy 666 GYNLIMNSEQPOGEGVTIAPRLITTVKQGLSTQSGI-----VPCWFLKEMLP 720  
 Db 691 -----LNEIESRCEBYPLTRAFQOLISTVSESPNLGKGLRPPGDFPLQFLRDSVFL 745  
 Qy 721 SYHKRKYNSHVREOJGCLILHAIIMLCHETDLSHTPEL-----QFLCIS--L 772  
 Db 746 RRTTRAPRAERKWEVAEVLLEVFLKLR-----DIEPQLEDFVQFELQSEETI 796  
 Qy 772 AYTBAG-----QVINIMGVDTIIDWMAAOPRSDAEGGQGLIKTV 818  
 Db 797 AKRPGSLMTHLINESPMLELALSLIESGVKOLD-TYAPFPGKLEKAVGHCLALML 855  
 Qy 819 KL-ASVTNNVTLRKPPSNVVSPLQAL---SQGHAGNNLJAVLAKYIYHKHPALPRL 874  
 Db 856 TLQKENVLFDLLRESQALIVCPLEQLQGINERTKADNVNMI-ARYLYHGN--TNPEL 912  
 Qy 875 AIQILKRLATVAPMS-----VYACIGNAAA--IRDAFLTL 909  
 Db 913 AFSAKILCCISCSNIOIKLVGDFTHDOSISQKMAGFVECLDCAEFPVLEBGSBL 972  
 Qy 910 OSKIEDMR--IKWILEPLTVAETOPGLIEF-LNLVYDGSOGSKESFLGMW---SC 962  
 Db 973 EKKLVAIRERTIHIHMLITSLDECPNIALYLLPELKKPVSTNLQDQGLGCPRTIC 1032  
 Qy 963 LHAVELIDSOODRYWCPPLHRAAIAFLHALMQ-----DRDSAMVLRTKPKF 1013  
 Db 1033 LHAIMILEKGTGR--TGPAVAVRESPOLAELCYOYIYOLCACSIDSGPTMRILRTSQP 1090  
 Qy 1014 WENLTPLEFGLTS--PSETS--EPSLLETICALIMKTIICLEIYVVGSLDOSIKOT--- 1066  
 Db 1091 -----LFSQLOLYLFPNSKXEYISMLNQMWSIMLKASIELRVT---SLNRQSHYORLL 1140  
 Qy 1067 -----LKFS-----TEKPAVMSGY----- 1082  
 Db 1141 HILLDMVPKPYSDGEGIEDENRVSQGFHPDTATKVRKIIIMLDSIDPSQELPEPQ 1200  
 Qy 1083 -----YKSL-AVHVAETEG-----SCT 1099

Db 1201 LDFFPRAQIEQYIANCEHNLNGQTVANVKLHRYLVAEVNALQGMALIGORELMEBIS 1260  
 Qy 1100 SILEY-----QMLVSARML--LIIATTHADIMHLD-SVVRQOLPVDY-- 1140  
 Db 1261 TVLQYVGNKLLQCLHARGHLESRLVETILTRCPDOLLQABERQOLIINDIODVHD 1320  
 Qy 1141 --LDGKALLVPASVNCILRSGMKCTLLILILARQKRELGSVDIELG----- 1187  
 Db 1321 KLIDBAQELMPVAGA--VFTLTRHLSQNVLTGKQ-----TSYLGSRABHAYAMLDS 1373  
 Qy 1188 -----LVEILSGVLQADQOLMEKTKAKYPSAFITVLOK 1221  
 Db 1374 CFTSPREBNPLVGFASIGDSLYILIKLDPILIKTGCG--PQRYVTHLYGSLLYLQIA 1432  
 Qy 1222 E-----MKVSDIQYQVLV--NVCEITQREVALPQPTHSLALGSATD---KDSM 1269  
 Db 1433 QRPDEBDTLEAKKTMERLTPREVPFSLQRENAITIE-----SYGALMEVCRDNC 1486  
 Qy 1270 ETDGSRSHRDQRDQVCVLGHLAKELCEVEDDSDWLQVTRRLPILPILLTLEVSIR 1329  
 Db 1487 DGHETGR-----MLALALDRIVSVQK--QOMLLYNSGYLVKIVDSLVBDR 1534  
 Qy 1330 MKON-----LHFTBATLHLLTLARTQGANVAVAGITQSLCLPLSVYQIST 1378  
 Db 1535 TLQSLTPQPLKALTYTESMAFLTRVAKIQGALBELRSGVI--VELAQCOVYDMKP 1592  
 Qy 1379 NGTAQTPASR--KSLDAPSWPGVYRLSMSLQGLKTRRYNLPBALDPVG-----VHQ 1431  
 Db 1593 ETDQSGMGKRPMPFIFPVDRYQOILPALQOCVILTSMAQLQAAQVLOPLSH 1652  
 Qy 1432 ERTLOCLNAVITVQSLACLBEADHTVGFILQLSNFKEMHFLPOLMR--DIQVNLGYLC 1489  
 Db 1653 SDTIALRQDOV--SAGSLQELALTLGISKAA-----LPLISLBDVNGSLM 1702  
 Qy 1490 Q-----ACTSLH--SRKMLQHYLONKQGDGIPSAVQVORPPSAASAAPS 1535  
 Db 1703 ELQGHIGRQRCIGLRSFGSDRLRQFKQDDVNEG-----DRV 1743  
 Qy 1536 SKOPADTASRQALHTVQYGLKILSKTLAALHFPDVQOILLDOSLDAEYFLPA 1595  
 Db 1744 SKKDEIEL-AMQOICANMEY-----COSLMDS--SPITQJA 1778  
 Qy 1596 ISFTTPTPDSV-----APSGTLATVNVALN-----MLG 1626  
 Db 1779 VCLFPPSLSETNRRQPRQDTQAPVVPYWRPLGDIITYLKQNSMDPSYDSHRQVS 1838  
 Qy 1627 EL-----DKKEPLQAV-----GLSTQAB-----GRTTKSLMFTM 1659  
 Db 1839 KLQNVBOUPPDEIKE-LQOSVAPAGVDKISTQKQVLAARRLVKVINNRKLLSLCSFTI 1897  
 Qy 1660 ENCFYLLISQANRYLADPAVH--PRDKQRMQELSELSTLSLRYFRGAPSPATG 1717  
 Db 1898 ETCPLTL-----WRHLEYTLHCPPTDQ-----DSLASKTLFSSRLQDSFAS- 1942  
 Qy 1718 VLPSPQKSTSL---SKASPSQEBLQI---VOAFVHMR 1753  
 Db 1943 -----ETNIDPRSGLAIVSQHLDQADAINAFBSLQK 1977  
 RESULT 16  
 08666  
 ID 08666 PRELIMINARY, PRT: 1139 AA.  
 AC 08666;  
 DT 01-UN-2002 (TRENBLrel. 21, Created)  
 DT 01-UN-2002 (TRENBLrel. 21, last sequence update)  
 DT 05-JUN-2004 (TRENBLrel. 27, last annotation update)  
 DE Hypochemical protein OSJNB0034B05.27;  
 GN ORFName=OSJNB0034B05.27;  
 OS Oryza sativa (japonica cultivar-group).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliopsida; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzae; Oryza.  
 OC NCBI\_TaxId=39947;

```

RN [1]
RP SEQUENCE FROM N.A.
RA Ming R.A., Yu Y., Soderlund C., Chen M., Kim H.-R., Rambo T.,
RA Saeki C., Henry D., Oates R., Simmons J.,
RA Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
RL [2]
RP SEQUENCE FROM N.A.
RA The Rice Chromosome 10 Sequencing Consortium;
RA "In-depth view of structure, activity, and evolution of rice
RT chromosome 10."
RL Science 300:1566-1569 (2003).
RN [3]
RP SEQUENCE FROM N.A.
RA Buell C.R., Wang R.A., McCombie W.R., Messing J., Yuan Q.;
RA Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC097446; AAM14698.1; -
DR EMBL; AE017058; AAP52188.1; -
DR Gramene; O86666; -
KW Hypothetical protein.
SQ SEQUENCE 1139 AA; 127147 MW; 7F50B31552B999FD CRC64;

Query Match 3.2%; Score 292; DB 2; Length 1139;
Best Local Similarity 18.7%; Pred. No. 4e-10;
Matches 226; Conservative 220; Mismatches 414; Indels 346; Gaps 56;

OY 25 LGRSLRLREISQIEA--ELANKMRRLLEGLSTYKPPSPSSAEKVKANKQVAA----- 73
DB 45 LDRALAAPSPVDALAEIRIKAHAMLRGVSMPVKPSDASRGALDSRYVVGSEHRLAVKP 104
OY 74 PLKELGLRISKEFLGDEEQSVOLLQCYLQEDYRGCTDSVKTVYADERSOALILKLADYX 133
DB 105 ELKAAALRLSKCMNDEVOISYLV-----KRTSENPFTALVAD--TEEFRLAVSVQY 154
OY 134 YEEERTCLRCVLAHLTYFQDERHPYRVEXADCVKLKELVSKYRQOFEE----LYKTE 188
DB 155 YLERQCLAKCIRIRIFVHAND-----CSDSIDAVREASVLYVREVEQRLLSIVRDEL 206
OY 189 APTWETHGMLTEROVSRFVQCLAREQSMLEIFLYIYAFEMAPSD---LVLVTMPFE 245
DB 207 ASAFVVKGG--AELTIS-WLEBTLIRINIFILFLFF-YDMSRNCGLMIMLCISIFXD 262
OY 246 QGFGSRQOTRHLVD-ETMDPFVDRIGYFSA--LIVVEGMDIESLHKALDD--RREHL 298
DB 263 MLSSGYDVGKFAVSEAKNSF---HYAAQQLFILIQTLDESLKMRVDEVPBSGYS 318
OY 299 QPAQDGLICQMDCLMTFGDIPHNAPVLIAMALNHTLNPEBTSVVKIGGTALQLVN 358
DB 319 TFSVVDILEMDVSVKLPFAVAVESGPLLIAMAVF-----LCLVMSLPSGNNLGP 369
OY 359 FOYLTRLLQSLASGNDCTTSTACMCVGLLSFVLTSLLEHLTLGNQODIITPACVLA 418
DB 370 VSGFRGILRTFTIS-----AFVASY----- 388
OY 419 SLPELFWGTEPTSGIILDSVCGMFPHLSPILDL--LRALVSGSTAKKYSFLDKMS 476
DB 389 ---EISYQTEDESS-LGMIINLICEVVDGSESLCMQWDXSFIDGIR---FNHYERKN 440
OY 477 FYNELY--GKHPHDVISHEDGTLMRQOTPKLYPLGQTNLRIPQTVGQVM--LDDRAY 532
DB 441 GVTTLVAVRSDTDNNVNDQI--ELHSPISIFGIGCTT--IPGSHGIYILKVLDDVA 495
OY 533 LVRMWYSVSSWTLFTCEIEMLLHVSTADVIQHQVRKPIDLVHKVISTDLSADCLLP 592
DB 456 LVRMW-----DLCLALHADKSLAVQASQN---LGVIDGVRIDIAKIFC--- 537
OY 533 ITRIRYMLQRLTT--VISPPVDVIVSCNCLTVLAARPAKWTDLRHTGFLPFAHVP 650
DB 538 ---TSIFKYVEDFNNAVCWSKTGLMLAEMLSCV---PYHFNVALDCGF--FTQ-- 584
OY 651 SLSLQMTISAEGNAGGYGNILMNSQPOGEVGTIAFLRLITTLVKGOL-----GSTQSG 705
DB 565 ---SGVAASDMLSLGSLARMLPATSDSDSCS-----SLTTTVLDPALQVAKGAAD 635

```

```

OY 706 GLVPCVMFLKEMLPSSYHKWRYNH-----GVREQICLLILET-- 743
DB 636 IISSEIISVQYIIMNMHMKTKYSRMKTLTKVFDVYSCIQVSPFSKLGIIWEILL 695
OY 744 ----IHAIL--NLCHETDL---HSHTPSLQFLCISLAYEAGQTVINIGIVDTID 793
DB 696 YDSSISYVLIHLSMSTQLHSHSGSYCHDLDI-----EDIQVLVCCGPDIVF 744
OY 794 MYMAAOPSDGAEGGCGOGLIKTV--KLASVTNNVIRLKPNSVNSVLEQALQGHM 851
DB 745 YMLSNLP-----EREKSKRMRLRVGSKAFYGFVVLVKIA-----FRH 786
OY 852 GNNLAVTAKY-----IYKHPDAPRLAIQLKRLATVAPMSVYACLDND--- 897
DB 787 SCNIIDVLSERQPSIEMRTCIIFH-----LQAPFKILS---QSHVACSSEDN 834
OY 898 -----AAA-----IRDAFLRLQSG- 911
DB 835 RTSNKPWPTDNTDIFLKAIVAAAVFSLCETAYAKAQPOLMENNVPVNGSEIWRLOT 894
OY 912 -----KIEDRIKYMILEFLTAVETDP-----GLIELFLMLEYKQDS----- 949
DB 895 ISCIDDEVQXNEVVAIFMLLSAARYQADVALVWLRLVPLPSQSTRQAQSDNSAH 954
OY 950 -DGSKEPSLGWMSCLHVLIEL-----DSQQQDRVWCPPLHRAAIAFLHLMODRDS 1002
DB 955 EOSSKTFVLPNGSGNPRVLBOILGYTGSTELMDR---SPSLISGVLDLKLMBESGAQ 1011
OY 1003 AMLV--LRTKPKFMENTLSPFLGTU--SPPSET-----SEPSILLETCL--IMKICLEIYY 1053
DB 1012 IYLEKMSRFFWENLSCCIIAFAASYDIDSEVTEDEKSLRYCCLGTIFIMSYELP- 1070
OY 1054 VYKGS 1059
DB 1071 -LQGR 1075

RESULT 17
OQ 06PDG0 PRELIMINARY; PRT; 2003 AA.
AC 06PDG0;
DT 05-JUL-2004 (TRENBLREL. 27, Created)
DT 05-JUL-2004 (TRENBLREL. 27, Last sequence update)
DT 05-JUL-2004 (TRENBLREL. 27, Last annotation update)
DE Nup205 protein (Fragment).
GN Name=Nup205;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RC MEDLINE=2238257; PubMed=12477932;
RA Straubeberg R.L., Pelngold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Schermer C.F., Bat N.K.,
RA Altschul S.F., Zeeberg B., Burow K.H., Scheffer C.F., Hsieh F.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ueda T.B., Toshiyuki S., Carinoci P., Prange C.,
RA Rana S.S., Loggellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McGowan P.J., McKernan K.U., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren B.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton B., Kerteman M., Madan A., Shevchenko Y., Bouffard G.G.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Tuchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Gilmwood J., Schmutz J., Myers R.W., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smalins D.E., Scherch A., Schein J.B.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences."

```



RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
 RN (2)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J TISSUE=Brain;  
 RA Strainsberg R.;  
 RL Submitted (SEP-2003) to the EMBL/GenBank/DBJ database.  
 DR EMBL, BC058729; AAH58729.1; -  
 FT NON TER  
 SQ SEQUENCE 2003 AA; 226929 MW; A130BC7AE0E53D27 CRC64;  
  
 Query Match 3.2%; Score 291.5; DB 2; Length 2003;  
 Best Local Similarity 19.2%; Pred. No. 1e-09;  
 Matches 354; Conservative 250; Mismatches 657; Indels 583; Gaps 83;  
  
 7 TNSVSPGSSRREIWTIL---LGRSALRELSQIEALNKMRRLLEGSLYK--PPSPSS 61  
 2 VNSAASLWGPYKOIWTGVSALMRROPEAVHLMILKCH--KDPFISLPKPPKPVQOO 58  
 62 AEKV-KANDVASPLKEIGLR-----ISKFLGDEEOSYQILQC--YIQEDYR 106  
 59 HEKIKASSEGAVALGQGGTRLLPEGLIKEAPFISLPDIGELSANVELLAGHQPHRP 118  
 107 G-TRDSVKTVDODEROSQALLKIDYEEBRTCIKCVLHLLTFQDER-HPRVRYAD 164  
 119 GLTRGLVAVTL-----YMDGKRCIAN--SLRTLLQSHRGKTWLE--- 156  
 165 CVDKLEKELYSKYRQOFBEELYKTEAPTWETHGNLMEROVSRKPFVOCLEQSMLEIFL 224  
 157 ---LSPFLVS-----MTR-----FIDELMEQGLTYKVLT 184  
 225 YVAEYEMAPSDLLV---LTKMFEQGFSGRQTRNRLVD-----ETM----- 262  
 165 L-----SQIDVNNPEKLRERGLSEKHKRKEVSDLIKCRGSIASLPRAMQOSPL 236  
 263 --DPRVDKRGYSALIVEGMIDISLAKCALDRRELHQPADGLICOD--MCLMTEF 317  
 237 PKDDTLILGLHERVTEANGSIDAVNLCL-----MALYYCFPTSFIDOSTEER 286  
 318 GDIPHPAPVL-----LAWALRLHLLN--PE-----E 341  
 267 DMHILHPLTERQYVSTHSRLQDSQPKLPGLQATVRLAWALRLRGISQLPDTALAE 346  
 342 TSSVVRKIGGTALQIANVFOYTLRLQSLASGNDCTTACMCVYGLSFLVT--SLBIA 399  
 347 FTEADBAIAEAIADNVFLFS--EAVVLAENFQEBEYIRIRHSLIDFLAFPMKVK 403  
 400 TLGNQOD---IIDPACVLDADPSLP-----ELFWGTE 428  
 404 QLKNADEADARMIHMSIQMGNEPISLRRLDEHMLLIGELYKKNPFIHELALBYWCPSE 463  
 429 FT-----SGLG-----IIDSVCGMFPHLS-----PLIQLLRALVSGSTAK 466  
 464 PLQGTWGSYIGVAAHQRPQROGVLSKVRQMGDLRPITYIFLPLKGLANPQCAH 523  
 467 KYVSELDKMSFYNELYKHKPHDVISHEDGTLWRQTPKLLYPLGGQTNLRIPQGTVGVM 526  
 524 YCFSLI-----KVNSSHVENIQAGGSP- 547  
 527 LDDRXYLVWMEYSVSWTLFTCEIEMLHVSTADVICHQCKVKPILIDLVHNVISTDSI 586  
 548 -----VSWEPFHSLLYH--BHLRKDLPASAVQY--BHLPSRGITQK--BODGLI 593  
 587 ADCLLPITSRIYMLQ--RLTTVISP--PVDVIASVNC-----LTVLAA--RN 629  
 594 A--PLQGTSTITWENARLALCEHPQWTPVVIIIGLQCSIPVLKAEKLTLLAARFS 651  
 630 P--AKVWTDLRHGFPLPVAHPVSSLSQMSIABEGNAGYGNLLMNSQPOGEXVTIA 686  
 652 PEIAASLWOSLEYQILQTVRVPSQRAIGIEVE-----LNEIESRCBEYPLIRA 701  
 687 FLRLITLYKQGLGTSQGL-----VPCVMVLKEMLPSTYKTKYNSHGREQJGCLIL 741  
 702 FCQLISTIVSSFPFNLAGLNPGRPDYLPQLRDVSFLRFRTAYRRAAEKMEVAVYVL 761

QY 742 ELIHAIINLCHETDLSHSTBSL-----QFLCICS---LAYTAG-----Q 779  
 DB 762 EVFYTLR-----DYEQLDEDFVDQFVELQGEKRIIAYKPPGFSIMYHLLNSPMLR 812  
 QY 780 TYININGIGVDPIIDVMAAOPESDGAEGOGQOLLIKTYKVL-AFAYTNVVRILKPPSNV 838  
 DB 813 LALSILBSGVQLD--TYAPPPQKGLBEKAVQHCIALNMLTLOKENLFPDLARBSQALIV 871  
 QY 839 SPLBQAL---SOGHAGNNLIVAKTYIYHKIDPALPRIALQOLRLATVAPMS----- 889  
 DB 872 SPLBQLQGINPRTKKADNVNVI-ARYLYHGNN--PEIAFESAKILCICISGNSIQVM 928  
 QY 890 -----YVACIANDAA--IRDAFLTLQSGKIDMR--IKMIIIEPLTV 928  
 DB 929 VQDFTHDSQVSKLMAGVECELDYEDTEEFVAVESGELEKGLAARIHSTRHILNLLIT 988  
 QY 929 AVETOPGLIELF-LNLBYKDSGDSKESFLGMW-----SCLHAVELIISQOQDR----- 977  
 DB 969 SLERNPMLALYLLGFELKPIISTNNLDPPVLAGCPRTCLHAILNILEKGTGRGDPVAV 1048  
 QY 978 --YWCPRLLHRAALAFHAL--WQDRDSAMLVLTREKFMENLTSPFGLTSP--PSE 1030  
 DB 1049 REY---PQLADLCYQVITQLCACSPISGPTWRYRTSODF-----LFSQOLHLPFSNK 1098  
 QY 1031 TSPSILEFTCALIMKIICLETIYVYVKGSLDQSLQTLKKPSI-----EKPRAYMSGYK 1084  
 DB 1099 EHEISMLSQMSWMLKTSIEIRVT--SLNQRSHQRLHLLDMDVFKPSIDSGEGW- 1154  
 QY 1085 SLAVHVAETEGSSCTSLLEYQMLVSAMWMLIITTHADIMHLDQSV-----VRRQLFLD 1139  
 DB 1155 -----EDENRSYSGFLHPTATATVRKRIISI-----LDSIDFSGEIPRLQID 1197  
 QY 1140 VLDGKALLVPAVYNCIR--LGSMTLLIILRQKRELGSVDLGLPTEILBEVL 1196  
 DB 1198 FEDRAQBOVI---ANCEHKNLQGGTYVNCVTL-----HRLVAVRNALQWVA 1242  
 QY 1197 QADQOLEKTKAKVFSAPIT---VLQWKEKVSIDIPQYSQVLYNVCSTLOBEVIALPQ 1252  
 DB 1243 AIGQRPMLBEISTILQYVGRANKLOCTIAKRAHBSKQVLVILITACPOBELIQABDR 1302  
 QY 1253 TRHSIATGASATEDKDSMETDCCSRHRHDOBDGVCLGMLAKELCEVDEBDGDSWLQVTR 1312  
 DB 1303 ---QILINDLQDVHDKVLDD-----BAQSL----- 1326  
 QY 1313 RLPIPLITTLTLEYSLSKQKYLHTEBATHLLTLARQOQATVAVANGITQISCLPLIS 1372  
 DB 1337 MPVAVAGAVFTL-----TAHLISQA-----VRTBORQPLVSGGGAQ-----Y 1362  
 QY 1373 VYOLSTNGTAQTPASR-----KSLD--APSMGVYRLMSIMBOLK 1413  
 DB 1363 AFMDSLSITSSPAABESRVRGASIGDSSHLIILKKLIDFLIKTGGPQVRVATHLYGSILY 1422  
 QY 1414 TLRTNPLPEALDVFVGHQERTLQCLANAVRYOS-----LACLEBAD 1454  
 DB 1423 YLQIAQRPDEBDTEAAKKTWMBERTAPBEDVFSKLORENMAIISYGAALMEVUCRDACD 1482  
 QY 1455 -HTVGFILQS-----NFMKEMHPLPQAMRDIONGLGYLQOACTSLHSRMLQHY 1505  
 DB 1483 GHEIGRMALALDLRIYSVDQHQOMLYLS-----NSGITYLVLDVSLVDDDRITQSL 1534  
 QY 1506 LONKRGDGLPSAVQAVQVRPSAASAAAPSSKQPADTEASEQ 1549  
 DB 1535 L-----TPQPLRLKALTYBSKMAFLTRVAKEQ 1563  
  
 RESULT 18  
 AAH58729  
 ID AAH58729 PRELIMINARY; PRT; 2003 AA.  
 AC AAH58729;  
 DT 02-MAR-2004 (Tribble, 27, Created)  
 DT 02-MAR-2004 (Tribble, 27, Last sequence update)  
 DT 02-MAR-2004 (Tribble, 27, Last annotation update)



DE NUP205 protein (fragment).  
 GN NUP205.  
 OC Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxId=10090;  
 RN [1]  
 SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6; TISSUE=Brain;  
 RX MEDLINE=22386257; PubMed=12477932;  
 RA Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,  
 RA Altschuler S.F., Zeeberg B., Buecaw K.H., Scheffer C.F., Bat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Dichtenko L., Marubina K., Farmer A.A., Rubin G.W., Hong L.,  
 RA Stempleton M., Soares M.B., Bonaldo M.P., Casavant T.L., Schetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carinci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
 RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kesteman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Greenwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
 RA Kravynski M.I., Skalska U., Smallus D.E., Scherch A., Schein J.E.,  
 RA Jones S.J., Marra M.A.,  
 RA "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences."  
 RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RL [2]  
 RN SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6; TISSUE=Brain;  
 RA Strauberg R.,  
 RL Submitted (SEP-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: BC058729; AAHS8729.1; -  
 FT NON\_TER 1  
 SQ SEQUENCE 2003 AA; 226929 MW; A130BC7AE053D27 CRC64;  
 Query Match 3.2%; Score 291.5; DB 2; Length 2003;  
 Best Local Similarity 19.2%; Pred. No. 1e-09;  
 Matches 354; Conservative 250; Mismatches 657; Indels 583; Gaps 83;  
 QY 7 TTSVLSFRRSRRLTIL--LGRSALRELQIEELANKMRRLLBGLSYR--PPSSSS 61  
 DB 2 VNSAASLTGPGYDIDQWGSALMRQPEBAVHLLDMILKGG--KDFISLFGNPPKXVQ 58  
 QY 62 AEKV-KANKDVASPLKEGLR-----ISKFLGLDEGVSOLLQC--YIOEYR 106  
 DB 59 HEKIQKASSEGVAIGQGGTLLPQOLKEAFISDLFDIGELSAVELLAGEHQOPHP 118  
 QY 107 G-TRDSVTVLQDEROSQALLIKADYYEERTCILRCVLLHLLTFQDER-HPYEVAVD 164  
 DB 119 GLTRGLVAVLL-----VWDGKRCTAN--SLRTLIQSRGKTWLE--- 156  
 QY 165 CVDKLEKELVSKRQOFBELVTEAPWETHGNLATEROVSRWFOCLREQSMLEIITL 224  
 DB 157 ----LSPELVS-----MTTR-----FIDELMEQGLTYKVTL 184  
 QY 225 YVAYEMAPSDLV---LTKMKEOGFGSRQTNHLYD-----ETM----- 262  
 DB 185 L-----SQDVNNEPEKQREGRGLGSEKRRKEVSDIKECROSLASLAPAMACOSPL 236  
 QY 263 --DFPVDRIQVSALILVEGMDIESLHKCALDRREILHQADGLICOD---MDCIMLTF 317  
 DB 237 PKDITLLILIGHRTVETANGSLDAVNCLL-----MALYICPTSTIIDOSTER 286  
 QY 318 GDIPHHAVL-----LANALLRHTLN--PE-----E 341  
 DB 287 DDMIHLLPLTERQVSTIHSRLQDSQPKLPGLGATVLAVALALRGISQPDVYALAE 346  
 QY 342 TSSVVRKIGGTATQIANVOYLTRLLOSASGNDCTSTACMCVYGLSPVLT--SLERH 399

DB 347 FTEADEAIAELIADNVFLPLS---EAVVLAENFYOEFTYRIRIRIHLITDFLAFMPIYK 403  
 QY 400 TLGNQOD-----ITPACGVADLPSP-----ELFMQTE 428  
 DB 404 QLNKRDADAMIMHSIQMGNEPPLISLRDLEHMLLIGELYKKNPFLALEFWCSB 463  
 QY 429 FT-----SGIG-----IILDSVCGMFPLLS-----PLLQILRALVSGKSTAK 466  
 DB 464 PLQPTTWGSLTGVANQRPQROVVLKSPVQMGDLPLPTIYIPLTKLQGLANPQCH 523  
 QY 467 KYVSEFLDKMSFYNELYKRPDVISHEBDGLMRQFTPLLYPLGQTNLRIPQGTGVGM 526  
 DB 524 YCFSL-----KXNSSHVENIQAGGSP- 547  
 QY 527 LDDRAVLYRWEYSYSWTLFTCEIEMLHAVSTADVIGHCQRYKPIDLVKIVISTDSI 586  
 DB 548 -----VSMHFPSLILLYH--EHLKRDLPASADVQY--RHLPSSRGITQK--BQDGLI 593  
 QY 587 ADCLLPITSRYMLLQ--RLTVVISP---PVDVLAACVNC-----LTVLAA--RN 629  
 DB 584 A--PLQITSTIITMSRNARLLYCHRPQTPVAVVILGLQCSIPVYKABLKTLLAPKGS 651  
 QY 630 P---AKWTDLRHTGFLPFPVAHPVSSLSQMSABGMNAGYGNLLMNSBQPOGEYVTA 686  
 DB 652 PEIAASLTQSLLEYQILQTVAVPSQRAIGIYB-----LNEIESRCEBYPLTRA 701  
 QY 687 FLRLITTVKQSLGSGTSGCL-----VPCWPFVKEMLPYHKKRYNSHGVRQIGCLIL 741  
 DB 702 FCOLISTIVBSSFPSNLGAGLRPGPFDEYLOFLRDSVLEFRTRAYRBAEKMEVAEAVL 761  
 QY 742 ELIHAILNLCHBTDLHSHTPSL-----QPLCS--LAYTAG-----Q 779  
 DB 762 EFTYGLR-----DYEPQLEDVQDFVLOGBEELIAYKPPGFSMLTHLINESPMLE 812  
 QY 780 TVINIMIGIVDTIDVMAAOPRSDGAEQGGQGLIKTVKL-AFSTNNVIRLKPSSNV 838  
 DB 813 LALSLEEGVQGLD-TYAPPPGKHLKAVQHCALNLT/LQKENTLPMDLREBQALIV 871  
 QY 839 SPLQGLAL--SQGHAGNNLIYAVLAKTYHKHDPALPPLAIQLKRLATVAPMS----- 889  
 DB 872 SPLQGLQGINPRTKADNVANI-ARYLYHGNN--PELAESAKILICISNSNIQVM 928  
 QY 880 -----VYACIGNDAA--IRDAFLVTLQSLIEDMR--IKWMLDEPLTV 928  
 DB 929 VQDFTHDOSVQKLMAGVBCIDYEDTETFEYRVEGSELEKKAIRIRTRHILNLIT 988  
 QY 929 AVETQPLIEFL-INTLEVYKSDSGSKFSLGMW-----SCLHAVELIDSQODR----- 977  
 DB 989 SLERNPMLALYLLGFELKRP1STNNLQDPVLGCPRTCLAHILNILEKTEGRGDPVAV 1048  
 QY 978 --YWCPLIHRPAIAFLHAL--WQDRDSAMLVLTFRKFMENLTP1PGLTSP--PSE 1030  
 DB 1049 REY---POLADLCYOVIVOLACASDPTSGPTRKYATSDOF-----LFSQGLHLPFSNK 1098  
 QY 1031 TSEPSILFTCALIMKIICLEIYVVYKSLDQSLQTLKKFSI-----EKRPAYSGYVK 1084  
 DB 1099 EHEISLMSQMSWMLKTASTIERVT--SLNKORSSTQGLHLLLDMDVYKPSDQEGM- 1154  
 QY 1085 SLAVVAETEGSSCTSLLEYQMLVSAMPMLLIATTHADIMHLLTDSV-----VRQLFLD 1139  
 DB 1155 -----EDKRSVSGFLHFDATKVRKILSI-----LDSIDFSGQIPEPLQD 1197  
 QY 1140 VLDTGKALLLVPAVNCIR--LGSMTCTLLILLRQKRELGVSDEILGLPTEILBGLV 1196  
 DB 1198 FEDRAIQGVLI--ANCEKRLQGGTVCNVKLL-----HRVIVAEVNALQGMVA 1242  
 QY 1197 QADQGLMEKTRAKVSAFIT-----VLQKMKVSDIPQVSQVLNVVCTSLQSEVIALPQ 1252  
 DB 1243 AIGQRPILMEBISTTLQVVGANKLLQCLAHARHLESWRQVEITLTPCPBELIOAEBR 1302  
 QY 1253 TRHSIALSATBCKDSMETDQCSRSRHRDQDGVCLGLHLAKELCEVDEDSWLOVTR 1312  
 DB 1303 ---QLIINDLQDQVHDKVLDD-----EAAQEL----- 1326

QY 1313 RLPILPILTLTLEVSILRMKNLFTETLTLTLTARTOOGATRAVAGITGOSICLPLLS 1372  
DB 1327 -MEVAGAVFTL-----TALISQA-----VTEGROPLVSGPGEAQ-----Y 1362  
QY 1373 VYOLSTNGTQATPSASR-----KSLD--APSPGYYRLSMSGLEQLK 1413  
DB 1363 APFLDSLSITSSPAABSRPVPFASIGDLSLHIIKGLDPLFKTGGRQRTYLXGSLY 1422  
QY 1414 TLRINFLPEALDPVGVHQBETLLOCLNAVTVQS-----LACLEAD 1454  
DB 1423 YVQIARPPDEPDTLEAKKTWERTLTAPEDVFSKQRENNALIESYGALMEVVCSDACD 1482  
QY 1455 -HTVGFILDS-----NFMKEWHPLPQLMRDQVNLGYLCOACTSLHSRKLQRY 1505  
DB 1483 GHEIGWMLALALIDRLVSVDKQHOMLYLS-----NSGYLKLVLVDSLVDDRTQLSL 1534  
QY 1506 LQNKNGDLPSAVAGVORPPSPASAPSSSKQPADTEASBOQ 1549  
DB 1535 L-----TPQPLIKALYTESKAPFLTRAKEQ 1563

RESULT 19  
Q6ZOG1 PRELIMINARY; PRT; 2067 AA.  
AC 06ZOG1:  
DT 05-JUL-2004 (TEMBLrel. 27, Created)  
DT 05-JUL-2004 (TEMBLrel. 27, Last sequence update)  
DE 05-JUL-2004 (TEMBLrel. 27, Last annotation update)  
DN MKIA0225 Protein (Fragment).  
GN Name=MKIA0225;  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OC NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Embryonic tail;  
RX PubMed=14621295;  
RA Okazaki N., Kikuno R., Ohara R., Inamoto S., Koseki H., Hiraoka S.,  
Saga Y., Nagase T., Ohara O., Koga H.;  
RT "Prediction of the coding sequences of mouse homologues of KIAA gene:  
RT III. the complete nucleotide sequences of 500 mouse KIAA-homologous  
RT cDNAs identified by screening of terminal sequences of cDNA clones  
RT randomly sampled from size-fractionated libraries.";  
RL DNA Ref. 10:167-180(2003).  
DR EMBL, AK129093; BAC97903.1; -;  
FT NON-TER 1  
SQ SEQUENCE 2067 AA; 233812 MW; 84C09D69FD77B0E CRC64;

Query Match 3.2%; Score 289.5; DB 2; Length 2067;  
Best Local Similarity 19.2%; Pred. No. 1.5e-09;  
Matches 356; Conservative 244; Mismatches 645; Indels 609; Gaps 83;

QY 26 GRSALRELSQIEAEI-----NKWRRLLLEG-----LSY 53  
DB 53 GRSALSVLLQOGAGVGAASIMGPYDQIWQTVGSALMRQPEAVHLLDMILKKHKRDPISL 112  
QY 54 YK--PPSPSSAAKV-KANKDVASPLKEGLR-----ISKFLGDEEGSVQLIQ 98  
DB 113 FKNPKKNVQOHEKIKKASEGVAIOGQCTRLPPEOLIKAFIISLDLEPDISLAVELL 172  
QY 99 C--YVQEDYRG-TRDSVKTVLQDERGSQLIKIADYYEERTCILRCVHLILTYFOER 155  
DB 173 AGEHQQPHFPGILTRCLVAVL-----TWGKRCLAN--SLKTLIQRR 213  
QY 156 -HPYRVEYADVCKLEKELVSKYRQOFELYTEKPAFTWETHGNLWTERGVSWFVQCURE 214  
DB 214 GKTWTLF-----LSPELVS-----MTTR-----FTDELME 238  
QY 215 QSMLEIFLYLYAYFEMAPSDLLV--LTKVFKEGQSGRSQRTNRLVD-----ET 261  
DB 239 QGLTYKVLTL-----SQIDVNNRPFKQRRRGSLSEKIKRKEVSDLIKRCROSLAS 290

QY 262 M-----DPVDRIGYFSALLVGEKDISLAKCALDDBRRLHQPADGLICOD- 309  
DB 291 LRAMACQSPLPDQDTLLILGHLERYTVBANGSLDANVLCIL-----NALLYCPDT 340  
QY 310 --MDCLMLTRGDIPIHAPVL-----LAAALLRHTAN 338  
DB 341 SFIDSTEREDMWHPLPULTRQYVSTHSLQDSQPKWLGLOATVRLAVALALRGIS 400  
QY 339 --PE-----ETSSVVRKIGGTALIQLVNVPQYLTRLLQSLASGANDCTTSTACACVYGLSFP 391  
DB 401 QPDTYALAEFPADEBAIAELADNVLEFLS--BAVLAENFQGEYIRIRHSLIND 457  
QY 392 VLT--SLEHTYGNQD-----ITDACEVLADPSLP----- 421  
DB 456 FLAFPMRKVKQKLNKADBDARMINHSIQWNEPPISLRDLRHLMLIGELYKKNPFILE 517  
QY 422 ---ELFWGTRP-----SGLG-----IILDSVCGMF-----PHLSPLIQLR 456  
DB 518 IALBYWCSEBPLQPTWNGSTLGVANRPPQQRQVLSKFRQMGDLPLFTIIPYIKMLQ 577  
QY 457 ALVSGKSTAKRVYSEFLDKMSFYNELYKHPDVISHEDGLMRQTPKLYPLAGQTNIR 516  
DB 578 GLANGPQCAHYCFSL-----KVNSSHYE 602  
QY 517 IPQGVGVQMLDDRAVLYRWBYSTSSWTLFTCEIMLHVSTADVIHQORVKGITIDLV 576  
DB 603 NIQAGGSP-----VSWHEFFHSLILYH--BHLKRKDPSPADSVQY--RHLPSRGIT 649  
QY 577 HKVISTDISIADCLPITSRIYMLIQ--RLATVISP--PVDVIVASCNC----- 621  
DB 650 QK--QDGLIA--FLQISTITITWSEKARIALCHERPTQTPVVVILGLQCSIPPLKAE 705  
QY 622 LVTIAA--RNP--AKVWTLRHTEGFLPPVAVPVSSLSQMTSABGMNAGYGNLLMNSBQ 676  
DB 706 LKTLAFAKSPBIAASLWQSLLEYQILQTVRVPQRQAIQIEVE-----LNEIES 755  
QY 677 POGEGVTLAFLRLITTLVYKQLOGTOSQGL-----VCWAFVLKEMLPSTHKRYNSHG 731  
DB 756 RCEBETPLRACQLSTLVESPPSBNLGAIRPGPDFDYLOFLRDSVLFRTFTRAYRBA 815  
QY 732 VREQIGCLILHAIINILCHETDLSSHPSL-----QPLICG--LAATYEAQ----- 778  
DB 816 EKMEYAEVVLVFFYKLR-----DYEQLEBPVQFVBLQSEELIAYKPPSPSLMY 866  
QY 779 -----QVINIMIGVDITDMVMAQPRSDGAEQGGQGLIKTVKL-AFSVTNNV 828  
DB 867 HILNESPMLBIALSLIBEGVQLD--TYAPFPGKHLERKAVOHCLALINLTLQKENVL 925  
QY 829 IRLKPPSVVSLPQAL--SQHAGGNLAVLAKYIYHGHDPALPRLAIQOLKRLATV 885  
DB 926 LBSQGLATVSLPBLQGLQGINPRTKADNVNIT--ARYLYHNNN--PELAESAAXILCTI 982  
QY 886 APMS-----VYACLGNDAA--TRDAFLTLQSKIEDMR--I 918  
DB 983 SCNNIQVKNWGDPTHQSVQKIMAGVBECLDVBTEPFRVYBEGSLEKKLAIRRET 1042  
QY 919 KVMLEFLTAAVETQPLIEIF-LNLEYKQSDGSKESFSLGMW-----SCHAVLELDSQ 973  
DB 1043 RIHITINILITSLERNPMLAYLLGFELKPIISTNNQDPVGLCPRCTLAHILNILEKG 1102  
QY 974 QODR-----WCPPLRHAATAFLAH--WODRDSANLVLRKPKFENLTSPLFG 1023  
DB 1103 TEGRDGPVAVARE--POLADLCYQVILQACASDTSGETRYKRYLITSDP-----LFS 1152  
QY 1024 TLSP---PSRTSEPSIETCLIMKIICLEIYVYVKSGLDQSLKDTLKKFSI-----EK 1074  
DB 1153 QLGHLPPSNKKEHLSMLSQSWLKTASIELRYV--SARQSRHTQRLHLLDMDMPVK 1209  
QY 1075 RPAVWSGYVSLAVHVAETBSGCTSLLEYOMLVANRMLIATTHADIMHLDTSV-- 1131  
DB 1210 PYSDBEGGM-----EDKRRSVSGFLAHPDTATKVRKILISI-----LDSIDFS 1251

QY 1132 --VERQLFDVLDGKALLLVPAVNCNR---LGSMTCLLLILRQKRELGSDVDEILIG 1186  
 DB 1252 QEIBEPLOLDFPDRAQIEQV---ANCENKLOQGTVCNKL-----HRLV 1296  
 QY 1187 PLTEILBEGVADQOLMEKTKAVSFARIT---VLQNKEMKVSIPQYSQVILNVCETL 1242  
 DB 1297 AEVNALQGMALIGORFLMBEISTILQYVGNKLLQCLHAKRHALBESMRQVBEIILTAC 1356  
 QY 1243 QEBVIALFDQTRHSLALGSATEDKDSMETDCSRRHNDQRDGVVGLHLAKELCEVDE 1302  
 DB 1357 POELIOAEDR---OLIRDLLQDVADKVLDD-----EAAQEL----- 1390  
 QY 1303 DGDGMLQVTRRLPLPTLLTTEVLSRKONLHPTBATLHLLTLARTQOGATAVAGAGI 1362  
 DB 1391 -----MPVAGAVFTL-----TAHLSQA-----VTEBQRPVLSGFG 1423  
 QY 1363 TOSICLPULSVYOLSTNGTAQTPSAR-----KSLD--APSPGVYRL 1403  
 DB 1424 AQY-----AFMLDSSLTSPASBRPVGFASIGDSSLHILKLLDPFLKTTGGGFGQV 1476  
 QY 1404 SMSLMEQLLKTNRNPLPEALDFGVGHQERTLOCLNAVTVOS----- 1446  
 DB 1477 RTHYGLSLLYVLIQIAQRDEDEPTLEAAKTKWERTLTAPEVDVFSKLQRENMALIBSYGAL 1536  
 QY 1447 --LACLEBAD-HITVGFILQLS-----NPKKHHFLPLQMLRDIQVNLGLQACTSL 1495  
 DB 1537 MEVVCRAQCDGHEIGRMALALDLRIVSDVQKHQWLLTYS-----NSGLKTVLDSL 1588  
 QY 1496 LHSKMLQHYLQNKKGDLPSAVAQRVORPSPASASASSSKOPADTEASEEQ 1549  
 DB 1589 VDDDRITQSL-----TPQPLKALTYTESGMALITVAHEEQ 1627

## RESULT 20

BAC97903 PRELIMINARY, PRT, 2067 AA.  
 ID BAC97903  
 AC BAC97903  
 DT 02-MAR-2004 (Tremblrel. 27, Created)  
 DT 02-MAR-2004 (Tremblrel. 27, Last sequence update)  
 DT 02-MAR-2004 (Tremblrel. 27, Last annotation update)  
 DE MKIAA0225 protein (Fragment).  
 GN MKIAA0225  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 CX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Embryonic tail;  
 RA Okazaki N., Kikuno R., Ohara R., Inamoto S., Koseki H., Hirooka S.,  
 Saka Y., Nagase T., Ohara O., Koga H.;  
 RT "Prediction of the Coding Sequences of Mouse Homologues of KIA Gene:  
 RT Ii. The Complete Nucleotide Sequences of 500 Mouse KIAA-homologues  
 RT cDNAs Identified by Screening of Terminal Sequences of cDNA Clones  
 RT Randomly Sampled from Size-fractionated Libraries.";  
 RL DNA Res. 10:167-180(2003).  
 DR EMBL; AK129093; BAC97903.1;  
 FT NON TER 1  
 SQ SEQUENCE 2067 AA, 233812 MW, 84C09D69F6D77B0E CRC64;  
 Query Match 3.2%; Score 289.5; DB 2; Length 2067;  
 Best Local Similarity 19.2%; Pred. No. 1.5e-09;  
 Matches 356; Conservative 244; Mismatches 645; Indels 609; Gaps 83;  
 QY 26 GRSALRELSQIABE-----NKIMRRLBS-----LSY 53  
 DB 53 GRSALSVLLQGGAGVGAASLNGPYKDIQWVGSAALMRQPEAVVLLDLMLTKKHKDPISL 112  
 QY 54 YK--PPSPSABKV--KANKDVASPLKELGR-----ISKFLGDEBQSVQLQ 98  
 DB 113 FKNPKPVQGHKQKASBQVAILQGGQTRLLPQLKBAFIIISDLDIGLSAVELL 172  
 QY 99 C--YLQEDYRG--TRDSVXTVLQDERQSAALLIKADYTYERTCILRCVLLHITTFQDER 155

DB 173 AGEHQHPFGLTGLVAVLL-----YMDKRCIAN---SLRTLIQRR 213  
 QY 156 -HPYREYADCVDLKELVSKYRQGFEBELYKTAFTWETHGNLMTREQVRWFOCLRE 214  
 DB 214 GKWTLLR-----LSBELVS-----MTTR-----PTDELM 238  
 QY 215 QSMLEIIFLYVAFENAPSDLV---LTKMFKEQSGSRQTNHLYV-----ET 261  
 DB 239 QGLTYKVTLTL-----SQIDVNNBEFKQRERGASBKHKRVSDLIKECROSLAS 290  
 QY 262 M-----DEFVDRIGYFSAIIVEGMDISLKCALDRREHQAQDGLICOD- 309  
 DB 291 LFMAQSPKPKDITLIGHLEAVTYEANGSLDVAVNLCL-----MALLYCPDT 340  
 QY 310 -MDCLMTFQDIDPHAPVL-----LAMALIRTLN 338  
 DB 341 SFIDQTEERDMTHHPLFLTERQVSTHISRLQDSQPKLGLQATVRLAMALALRGIS 400  
 QY 339 --PE-----ETSSVVRKIGTAIQANVFOYTRILQSLASGNDCTSTACMKVYGLISF 391  
 DB 401 QLPVYTALEFTEADEALAEIALADNVFLFS---EAVVLAENFYQEBFYIRIRHSLITD 457  
 QY 392 VLT--SLELTIGNQOD---IIDACEVLADPSLP----- 421  
 DB 458 FLAMPKVKVQLKRADEADAMHMSIQMGNEPPISLRDEHMLLIGELYKKNPFILE 517  
 QY 422 --ELFWGTEPT-----SGLG-----IILDSVCGMF-----PHLSPILQILR 456  
 DB 518 LALEWCPSEPLQPTWNGSYLGAHQRPQQRVLLSKFVRQMGDLLEPTIYIPLYKQLQ 577  
 QY 457 ALVSGKSTAKKVVYFLDKMSFYNELYKHKPHDVLSHEDGLMRQTPRLYPLQGGTMR 516  
 DB 578 GLANGPQCAHYCFSL-----KVGSSVBE 602  
 QY 517 IPQGVQVMDDAVYIVRWEYSYSWTLFTCEIEMLHVSTADVIOHCORVXPIDLV 576  
 DB 603 NIQAGSGP-----VSWHFPHSLLYH---EHLKDLPSADSVQY--RHLPBQIT 649  
 QY 577 HKVISTDLISADCLPITSRIYMLQ--RLTTVISP--PVDVIASCVCN----- 621  
 DB 650 QK--EODGLIA--FLQLTSTIITWSENAKLCHHPQWTPVVLIGLLQCSIPLYKABL 705  
 QY 622 LTVLAA--RNP---AKWTDLRHGFLPFPVHPVSSLSQMTSAGMAGGYNLMSBQ 676  
 DB 706 LKTLAARFESPEIASIMQSLQYQILQTVAVPQROAIGIEV-----LNEIBS 755  
 QY 677 PQGEYGTAFRLRLITLVKQLGSGTOSQGL-----VPCVWFVLEKMLPSYHKVRVNSHG 731  
 DB 756 RCSEYPLTRACQLISTLVESSPSNIGAGLRPGFPDYLOFLDSDVLEFRTRATYRGA 815  
 QY 732 VREQIGLLELHAIALNLCHEIDLSHSTPSL-----QFLCIS--LAYTAG----- 778  
 DB 816 EKMEVAVVLEVPFKLR-----DYEQLDPDVQGFELQGBEIIAVKPGFSIMY 866  
 QY 779 -----QTVINMIGVDITIMMAAQRSQASQGGQGLIKTVK--ASVTNVV 828  
 DB 867 HLINESPMLBIALILBEGVQQLD--TYAFPQKHLBEAVQHCALNLTLOKXNLFMDL 925  
 QY 829 IRLKPSNVVPLQAL---SQHGAGNNLAVLAKYIYHGDALPLALQILKRLATV 885  
 DB 926 LRESQALIVSPQLQGINPRTKADNVNT--ARTIYHGNN--PELASAKILDCI 982  
 QY 886 APMS-----VYACIANDNAA--TRDAFLTLQKIDMR--I 918  
 DB 983 SCNNINIQVQWVGDPTHQSVSQKLAGFVELDVEDTEFPVRVEGSGELBKGLAIRBT 1042  
 QY 919 KVMILFETAVETOPGLISLP--LNLVYQSGDSGSKESLGMW-----SCHAVLELDSQ 973  
 DB 1043 RIHILNLTLSLBNPPLALYLLGFLKPKISTTNLQDPVLCGPRYCAHAIINLIEKG 1102  
 QY 974 QQDR-----YMCPLRHAAALAFHLA--WQDRDSAVLVLRTKFWMENLTSPLFG 1023

```

Db 1103 TEGRDGPAVAREY---POLADLCYQVITQACAGSDTSGPTWRYLRTSGDF-----LFS 1152
Qy 1024 TUSP---PSETPSPILFETCALIMKICELIYVVKSLDOSLDTAKKST-----EK 1074
Db 1153 QLOHLPSPNKEHISMUSQSMWMTKASISLRT---SLNRORSHQORLHLHLLDDMPVK 1209
Qy 1075 RFAVWSGVYKSLAVHVAETEGSSCTSLFQMLVSAVRMLLIATTHADIMHLLTDSV--- 1131
Db 1210 PYSDGSGGM-----EDENRSVSGFLHPTATKVRKRLST-----LDSIDBS 1251
Qy 1132 --VRQFLFDVLDGTALLVLPASVNCRL--LGSNKCTLLILLQMKRELGSVDIEIG 1186
Db 1252 QEIPEPLQDFPFRAGIEQVIT---ANCEHKLQGVTCVVKL-----HRVLY 1296
Qy 1187 PLTEIEGVLDQADQOLMEKKAIVPSAFIT---VLOMKMKVSDIPQSVLAVNCETL 1242
Db 1297 AERNALQGMALIQORPLMEISTIIQYVGNKLLQCHAKHAALESWQOLVEITLTC 1356
Qy 1243 QSEVIALPQTRHSLALGATBEDKSMETDSCSRSHRRQRDGVCVGLHLAKELCEVDS 1302
Db 1357 PQLIQAEDR---QLIIRDLQDVHDKVLD-----EAAQEL----- 1390
Qy 1303 DQDSMLQVTRRLPIELTLLTEFVSJRMKONLHFEATLHLTLTARTQOGATVAVAGAI 1362
Db 1391 ---MFWVAGAVFTL-----TAHLSQA-----VTRQORQPLVSGPGE 1423
Qy 1363 TOSICPLISVYQVSTNGTAAQTPASR-----KSLD--APSWGVVTL 1403
Db 1424 AQY-----AFMLDSSLTSSPAESRPVGFASIGDSSLHLLKLLDLTKTGQGFQRY 1476
Qy 1404 SMSLMEQLLKTLYNPLPALDFVGVHQRKTLQCLNAVTVOS----- 1446
Db 1477 RTHVLSLLYVLIQIAPRDEPDTLBAKKTWERLTPADVDVSKQOREMMAIIESVGA 1536
Qy 1447 --LACLEAD-HTVGITLQS-----NFMKEWHFHLPLQMRDIQVNLGYLCOACTSL 1495
Db 1537 MEVVRDADCDGHEIGMMLALDLRVSVDKQHWLTLYS-----NSGYLKVLVNLS 1588
Qy 1496 LHSRKLQHYLQNKQDGLPSAVAQVORPPSAASAPSSKQPADTASQ 1549
Db 1589 VDDRTQSL-----TPQPLKALTYTESKQALFTRVAKQ 1627

RESULT 21
O6P9L5 PRELIMINARY: PRT: 1851 AA.
AC O6P9L5:
DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
DE Nup205 Protein (Fragment).
GN Name=Nup205;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=1090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932;
RA Strauberg R.L., Feingold B.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Scheffer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Datschenko L., Marusina K., Farmer A.A., Rudin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.B.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carrinck P., Prange C.,
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Boeak S.A., McEwen P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villation D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Bailey J., Helton B., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,

```

```

RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.W., Butterfield Y.S.,
RA Krzywicki M.T., Skalska U., Smalios D.E., Scherch A., Schein J.B.,
RA Jones S.J., Maira M.A.,
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RA Strauberg R.L.
RN Submitted (NOV-2003) to the EMBL/GenBank/DBJ databases.
RP [3]
RC SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RA Strauberg R.L.
RN Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL, BC060711; AAH60711.1; -.
DR EMBL, BC060139; AAH60139.1; -.
FT NON_TER 1
SQ SEQUENCE 1851 AA; 209925 MW; DD11A9847159291 CRC64;

Query Match 3.2%; Score 286; DB 2; Length 1851,
Best Local Similarity 19.2%; Pred. No. 2.1e-09;
Matches 334; Conservative 236; Mismatches 621; Indels 550; Gaps 76;

Qy 15 RSSRLMTITLGRSLRLSOTIEALNKMRLL--BELSTYKPPSPSAEKYKAKOVA 72
Db 15 RFTBELMQGTLVYKTLTSLQI--DVNNEPRLQREKIG-----SEKIRKRS 61
Qy 73 SPLKELGRISKPL-----GLDEOSVOLLOCLQES--DNGTRDSVYTVLDERQS 122
Db 62 DLIRKCRSLAESLPANACOSPPLPRDYL-LIIGLEHYTVYANGSLDAV----- 110
Qy 123 QALIKADVYBERTCILRCVHL-L-TYFQDERHPYVEYADCVDKL-EKELYSKYRQ 180
Db 111 -----NLCLMALVLCPTSFIDQTEERDDMIHLLPLTERQYVSTHSR 156
Qy 181 FEELYKAPWETHNGMLNTERQVSRWFVQCLRQSMLEI----- 221
Db 157 LQ-----DSQPKPLPGIQAIVRLA--MAL-ALRGISQLPVTALAEFTADEAIAELA 208
Qy 222 --IFYAYVFEMADSDLLVTKMFQEGSGRQNRHLYDVTMPFVDRICVPALIVE 279
Db 209 DNVFLFL-----SEAVLAKENFYQSEFYIRRHSLITD-----FLAFM-- 246
Qy 280 GMDIESLHKCALDRLRHQPAQDG-----LICQMDCLMTFGDI-----PHNAPVLLA- 329
Db 247 PMVKQKLRADBDARMIHNSIQGNBPPISLRDLHMLLIGELYKQNPFLIELALEY 306
Qy 330 WALRHNLNPEISVYRKIGGTALQIANVOYLTRLLQSLASGANDCTSTACMCVYGL 389
Db 307 WC-----PSEPLQTPWGSYLGVA----- 326
Qy 390 SFVLTSLEHTLQAGQODIITPACVLAADPSLPFLFWGTEPTSGIILDSVCGMPHLLS 449
Db 327 -----HQRPRQVYLSKPYRQMGDLIP-----PTIYI 354
Qy 450 PVLQRLPALVSGSTAKVYSPFDKMSFYNELYKHKPHDYISHBDGLTMRQTKLHYPL 509
Db 355 PYLMLQGLANGPCAHYCSLL-----KV 379
Qy 510 GGCQNNALIPQGVQVWLDDBRATVYRVEYSWTLFTCEIMLHVSTADVIOCORV 569
Db 380 NGSSHVENIQAAGSP-----VSWNEHPSHLLTH-----EHLKDLPSADSVQY--RH 426
Qy 570 KPIIDLVAHYISTDLADCLPITSRIYMLQ--RLTVIISP--PVDYASVCNC-- 621
Db 427 LPSRGIQK--EDQGLIA--FLQITSTIITWSRNALALEHPQWTVVYILGLQCSIP 482
Qy 622 -----LTVIAA--RNP--AKVWTDLRRTGFLPFVAHPVSSLQNMISAEKNAGGYGN 669
Db 483 PVLKAEILKTLAFAFGSPBIAASIMQSLBAYTQILQVTRVPSQQAIGIVBE----- 533

```

```

QY 670 LMANSEPOGEGYVTAFLALITLVKQGLSTOSQGL-----VPCWVFLAKEMLPSTHK 724
DB 534 -LNEIEBRCSEYPLTRAFQCLITLVBSFSPNSGAGLRPGDPDYLOFLDLSEFLBRT 592
QY 725 WRVNSHVBRIOGGLIELIHAIIINLCHEITDLSHSHPSTL-----QFACISG---LAYTE 776
DB 593 RAYRRAAEKVEVAEVALEVFYKLR-----DYEQLQEDPDQFQELQGBEIIAYKP 643
QY 777 AG-----QVININGIGVDTIDMNAAPRSDAGGQGGOLLITVKL-A 821
DB 644 PGFSWVHNLNBSFMLELALSLBEGVQD-TYAPFGKGLKRAVCHALANTLQK 702
QY 822 FSTVNNVIRLKPNSVNSPLEQAL-----SQHAGNMLIAYLAKTYHGDALRLAIQL 878
DB 703 ENLFMDLIRBSQALIVSPLQLOGINPRTKADNVNI-ARYLYHNNN--PELAFES 759
QY 879 LKQLATVAPMS-----VYACIGNDAAA--IRDAFLTRLOSKI 913
DB 760 AKILCCISNSNIGVKNVGDFTHDQSVQKLMAGVECLDYEDTEBEFVREGESELEKCL 819
QY 914 EDNR--IKVMILEPLTVAVETOPGLIELF-LNLEVKDSGDSKEFSLGMW---SCLHAV 966
DB 820 AAIHETRIRHILNLTSLKBNPMLAYLLGFELKKRISTTNLODPGLCCPRTCLAI 879
QY 967 LELIDSOQODR-----YCPPLHRAALAFHAL--WDRRDSANLVIRTKPKFEMEN 1016
DB 880 LNIIEKTEBGRDGVAVAREY--POLADLCYOVYIOLCASCSDTSGPTRYLRTSDP--- 933
QY 1017 LITSLPGLTSP--PSETESESLIETCALINKIILEIYVYVKSGLSDLOTLKKFSI- 1072
DB 934 ---LFSOLOHLPPSNKEHEISMLSQSWMLKTSIEIRVT---SLNQRSHIOLHML 986
QY 1073 ---EKRFAYSGVYKSLAVHVAETEGSSCTSLLEYOMLVANEMLLIATTHADIMHL 1127
DB 987 LDDMPVYPSDGBGM-----EDENRSVSGFLHPTAKVVRKIKLSI----- 1028
QY 1128 TDSV-----VRQPLFLVDLGTKALLLVPAVNCIR--LSGKCTLLIILRQKREL 1179
DB 1029 LDISDFQGEIPEPLQDFFDRAQIEQVI---ANCEHKNLCQCTVCNVKL----- 1075
QY 1180 SVDEILGPTEILIGVLOADOQMLEKTKAYPSAIIIT---VLQKEMKVSDIPOYSOLV 1235
DB 1076 -HRLVAEVAVALQGMALIGORPLMEIISTILQYVGRNKLQCLAKRHALLESWRQV 1133
QY 1236 LNVCEITLOEVIYALPDOTRHSIALGASATEDKDSMETDCSRRHRDQDGVCLGLHAK 1295
DB 1134 EITITACQELIOADR---QLIRDLQDVHDKLDD-----EAYQ 1172
QY 1296 ELCEVDEBDSGLQVTRRLPILPTLTLEVSRLMKONILHTEATVILLTLARTQCAT 1355
DB 1173 EL-----MPVVAAGVFTL-----TAHLSQA-----VREIQROP 1200
QY 1356 AVAAGGITQICPLPSTVQSTNCTAOTPSASR-----KSLD--APS 1396
DB 1201 LVSPBGEAQ-----YAFMLDSSLTSSPAPASRNVGASIGDSSLMHILKKLDPFIKT 1253
QY 1397 WPGVYRLMSLMEOULTKTYRNFLEALDFVGVHDERTQCLINAVTVOS----- 1446
DB 1254 GGGFORVATHLYGSLLYLQIAQRPDEPDTLEAAKTKMERLTAEDVPFSLQREMAII 1313
QY 1447 ---LACLEBAD-HTVGFILQLS-----NFMKEWHFLLPOLMRDIQVNLGYL 1488
DB 1314 ESYGALLMEVVCRCADCGHEIGRMALALDRIVEVDKOHQWILYLS-----NSGYL 1365
QY 1489 COACTSLHRSKMLQHYIQNKNGDGLPSVAVQVRPSPASAAASSSGKOPAADTEASEQ 1548
DB 1366 KVLVDSTLVDDBRTQSLT-----TPQPLKALYTESKMAFLTRVAKQ 1410
QY 1549 Q 1549
DB 1411 Q 1411

```

```

RESULT 22
ID AAH60139 PRELIMINARY; PRT; 1851 AA.
AC AAH60139;
DT 02-MAR-2004 (Tremblrel, 27, Created)
DT 02-MAR-2004 (Tremblrel, 27, Last sequence update)
DT 02-MAR-2004 (Tremblrel, 27, Last annotation update)
DE NUP205 protein (Fragment).
GN NUP205.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N. A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RX MEDLINE=22386257; PubMed=12477932;
RA Strauberg R.L., Pelngold B.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Sherman C.M., Bhat N.K.,
RA Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Prange C.,
RA Brownstein M.J., Uedini T.B., Toehiyuki S., Carninci P.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.C., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green B.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butlerfield Y.S.,
RA Krzywinski M.I., Skalske U., Smalins D.B., Schermer A., Schein J.B.,
RA Jones S.J., Maira M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N. A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RA Strauberg R.;
RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL, BC060139; AAH60139.1; -.
FT NON-TER
SQ SEQUENCE 1851 AA; 209925 MW; DDF11A9847159291 CRC64;

Query Match 3.2%; Score 286; DB 2; Length 1851;
Best local similarity 19.2%; Pred. No. 2,1e-09;
Matches 334; Conservative 236; Mismatches 621; Indels 550; Gaps 76;

QY 15 RSSRBLWTILGSRALBELSQIBELNKHWRLL--EGLSTYKPPSPSAAKRVANKOVA 72
DB 15 RFTBELMEQGLTYKVLTLISQI--DVNNBEFKLRERGLG-----SEKIRKEVS 61
QY 73 SPLKELGIRIKEL-----GLDEBSVOLLOCCYGE--DYRGTRDSVYTVQDERQS 122
DB 62 DLIRECROSLAESIPFAMACGSPLEKDDTL-LHGLELVTVANGSLDAV----- 110
QY 123 QALILKADYAYBERTCILRCVLIHL--TYFODERHPYVVEYADCVKXL-EKELVSKYRQ 180
DB 111 -----NLCILMALVLCFDTSPFIDQSTEBRDMDIHHILPILTERQYVSTHSR 156
QY 181 FEEDYKTEAPTWETHGNLMTERQVSRMFVQCLREOSMLBT----- 221
DB 157 LQ-----DSQPMKLPGLOATVRLA--WAL-ALRGISQPLDYALAEFTDEADAJAELAIA 208
QY 222 --IPLYYVPEMASSDULVLTQMPKQGFSGRQTNRLHVDSTMPVDFVDRIGVFSLILVB 279
DB 209 DNVPLFTL-----SEAVVLAENFYQBEFYIRRHSLITD-----FLAFM--- 246
QY 280 GMDTSLHKCALDDBREHQAQDG-----LICQMDCLMTTFQDI--PHHAPVLLA- 329

```

247 PMKVKQLKNRADBDADAMHMSIQMGNEPPISLRDLBHLMLIBELKXNPFHELEALEY 306  
 QY 330 WALKRHTLNPEETSVAKIGGTALQNLVFOYLRLLQSLASGANDCTTACMCVYGLL 389  
 DB 307 WC-----PSEPIQOTPTVMGSYLGA----- 326  
 QY 390 SFVLTSLEHNLGNQODIIDFACVADPBLPELMGTEPTSGIGTILDSVCKMFHLLS 449  
 DB 327 -----HOPRQROVAVSKFVRQMGDLPL-----PTIYI 354  
 QY 450 PLLQLRALVSGSKTAKKVSFLDKMSFYNELYKHKPHDVI SHEDGTLMBRQTPKLLYPL 509  
 DB 355 PYLMLQGLANPGCAHYCSFL-----KV 379  
 QY 510 GCGONLAIPOGTGQWLDRAVLYVWEYSWTLFTCEIEMLHVSTADVIQRCQV 569  
 DB 380 NSSSHVNNIQAGGSP-----VSWHFPHSLTLTH-----EHLKOLPSADVQY--RH 426  
 QY 570 KPIIDVHKVISTDLSTADCLPITSRIMLQ--RTTVISF--PVDVIACVNC-- 621  
 DB 427 LPSRGITQK--EGDGLIA--FLOUTSTITWSENALALCEHPQWTPVVITGLQCSIP 482  
 QY 622 -----LTVLAA--RNP--AKWTDLRHTGFLPFVAHFVSSLSQMSAEGMAGYGN 669  
 DB 483 PVLKAEILKTLAFAKSPETIASLMOGLEVTQILQTVRVPDSQQAIGIEV----- 533  
 QY 670 LMKNSBQOGEVGTATPLITLTVKGOUGTOSQGL-----VPCMFVLKEMLSYHK 724  
 DB 534 -LNEISRCEBYPLTAPCOLISTLVESFSPNLGAGLPPGDFPLQGLRDSVFLRFT 592  
 QY 725 WRNSHGVBEQIGCLLELHIALNLCHETDLHSHTPL-----QFLCIS--LAAYE 776  
 DB 593 RAYRAAEKWEVAEVLVLFYKLLR-----DVEPQLEPVOQFVLOEHEIITAYP 643  
 QY 777 AG-----QVINIMGIVDTIDMMAAPSDAEGGCGQLIKTKVLA-A 821  
 DB 644 PGEISMTHLLNESPMLTELALSLIEBGKQID--TYAPFPCKHLEKAVQCHLALMLTLQK 702  
 QY 822 FSVTNVAVILKPPSNVVSPLLEQAL--SOHGANNNLILAVLAKYIYHKDPLPRLAQ 878  
 DB 703 ENLFPDLAESQALIVSPLLEQLOGINRTKADNVVNI--AKYIYHGNN--PELAFES 759  
 QY 879 LKRLATVAPMS-----VYACLGNDAAA--IRDAFLTLQSKI 913  
 DB 760 AKITLCISGNSNIQVAVMDFTHDQSVQKLMAGFEGLDVEDTEFVRVBSGLEKXL 819  
 QY 914 BDMR--IKMILEPLTVAVETOPGLIELF--LMEVYKDSGSKESLSGMW--SCLHAV 966  
 DB 820 AAIKRETRHILMLITSLERNPNVALYLLGFELEKPISTTNLODPVLCGPRCTLHAI 879  
 QY 967 LELIDSQOQDR-----YWCPRLLHRAALAFIAL--WODRDSAMLVLRTPKPMEN 1016  
 DB 880 LNLIEKTEGREGDPAVAREY--POLADLCYVITQLCASDTSOPTKRYRTSODF-- 933  
 QY 1017 LTPLEFGLTSP--PSETESESLTCLINKITILEIYVVKSLDLSKDTLKKESEI- 1072  
 DB 934 ----LPSQLOHLPSGNKEHEISMLSQMGMKMTASIEIRLV--SLNRQSHQRLHML 986  
 QY 1073 -----EKRFAYWSGVKSLAVHAVETEGSSCTSLLEYQMLVSNRMMLIITTTADIMHL 1127  
 DB 987 LDMVAVKYSDBEGM-----EDENRSVSGFHFDTAKVRKIKLISI----- 1028  
 QY 1128 TDSV-----VARQFLVDLDTKALLVPAVNCIR--LGMKCTLLILRLQMKBELG 1179  
 DB 1029 LDSIDFQOEIPEPQLDFFDAQIEOVI--ANCEHNLQOQYVQVYKLL----- 1075  
 QY 1180 SVBDEILGPIREIEGVQADQOLMEKTKAKVSAFIT--VLOMKEMVSDLPQYSQV 1235  
 DB 1076 --HRLVAEYVNAQGMALIGRPLIMEISTILQYVGRNQLQCLAKRAHIALESBQV 1133  
 QY 1236 LNVCEITQIEVIALFDQTRHIALGSAITEDKDSMETDCCSRRRDQDVCVLGHLAK 1295  
 DB 1134 EILITACPOEHLQAEHR--QLIRDLQDVHAKVLD-----EAAQ 1172

QY 1296 ELCEVEDGDSMLQVTRRLPIPLTTLTTFVSLRBMKONLHFEATLHLTLARQOGAT 1355  
 DB 1173 EL-----NPPVAGAVFTL-----TAHLSQA-----VREOROP 1200  
 QY 1356 AVAGAGITQSICLPILSVYOLSTNGTAQTPSASR-----KSLD--APS 1396  
 DB 1201 LVSGPGEAQ-----YAFMLDSBLTSSPAAESRVRVGPASTODSSLIHLIKKLDLPILKT 1253  
 QY 1397 WGVYRRLNSLSMLEQLKTLRNFPLBALDFVGHQERTLOCLINAVTVOS----- 1446  
 DB 1254 GGGFQRVATHLYGSLTYLTYLQIAQRPDEPTLEAAKMTWERTLTAPEDEVFSKLQRENMAII 1313  
 QY 1447 -----LACTERAD--HTVGFILQIS-----NPMKMHPLPOLAMDIONVGYL 1488  
 DB 1314 BSYGAALMEVCRDCCDCHIEGRMLALMLLRIVSVQKHQMLYLS-----NSGYL 1365  
 QY 1489 COACTSLHSRKMLOHTYQNTKNGDGLPSAVQVQRPSPASAAAPSSSKCPADTEASBQ 1548  
 DB 1366 KVLVDSLVDDBDRTILQSLT-----TPQPLKALVTSKMAFLTRVAKQ 1410  
 QY 1549 Q 1549  
 DB 1411 Q 1411

RESULT 23  
 AAH60711  
 ID AAH60711 PRELIMINARY, PRT, 1851 AA.  
 AC AAH60711;  
 DT 02-MAR-2004 (TREMblrel. 27, Created)  
 DT 02-MAR-2004 (TREMblrel. 27, last sequence update)  
 DE 02-MAR-2004 (TREMblrel. 27, last annotation update)  
 GN NUP205 protein (fragment).  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 CC NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6; TISSUE=Brain;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heide P.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stedelson M., Soares M.B., Bonaldo M.P., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Ueda T.B., Tohyiuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mallary S.J.,  
 RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richardson S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren B.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettelman A.C., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko I., Bouffard G.G.,  
 RA Blakeley K.W., Touchman J.W., Green B.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butlerfield Y.S.,  
 RA Krzywiński M.I., Skalska U., Smallus D.E., Scherch A., Schein J.E.,  
 RA Jones S.J., Marra M.A.,  
 RT Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences",  
 RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6; TISSUE=Brain;  
 RA Strausberg R.,  
 RA Submitted (NOV-2003) to the EMBL/GenBank/DBS databases.  
 RL EMBL, BC060711; AAH60711.1;  
 FT NON TER 1 209925 MW; DDF11A9847159291 CRC64;  
 SQ SEQUENCE 1851 AA; 209925 MW; DDF11A9847159291 CRC64;

Query Match 3.2%; Score 286; DB 2; Length 1851;

Beet Local Similarity 19.2%; Pred. No. 2.1e-09;  
Matches 334; Conservative 236; Mismatches 621; Indels 550; Gaps 76;

Qy 15 RSSEBMTLLGRSLRELSSQIBELNKGMRLL--EGLSTYKPPSPSSAEKRYANKOVA 72  
Db 15 RFTBELMEQGLTYKVLTLSSQI--DVNNEPEKLOREGIGL-----SEKRRKKS 61  
Qy 73 SPLKELGIRISKPL-----GLDEBSVOLLOCTLOE--DYRGTRDSVKTVALDEROS 122  
Db 62 DLIRECROSLAESLSPAMACOSPLFPDQTL-LIGHLEKRYEVAANSIDAV----- 110  
Qy 123 QALIKLADYYEERTCILRCVTLHL--TYFODERHPRYEYADCVDKL-EKELVSKYRQ 180  
Db 111 -----NLCILMALLYCFDTSFIDQSTEEBDDMIHMLPLTERGYVSTHSR 156  
Qy 181 FEELYKEAPFWETHGNLINTERQVSRWFVQCLREOSMLEI----- 221  
Db 157 LQ-----DSQPKLPEGLQATVLA--MAL-ALRGISQLPDVTLAETEADEAIAELA1A 208  
Qy 222 --ILYVAYFEMABSDLLVLTMPKQSGFSGRTNHLVDENDFVDRIGFSLIIVE 279  
Db 209 DNVEFLPL-----SEAVVLAENFYQEEFYIRIRHSLITD-----FLAFM-- 246  
Qy 280 GMDIESLHKALDDBRELHOFADG-----LICQMDCLMLTFGDI--PHHAFVLA- 329  
Db 247 PMKQKQKRNADDEBARMHMSIQNGNEPPISLRDLHMLLIGLYKKNPHELEALEY 306  
Qy 330 WALLRHTLNPETS SVRKIGTATQNLNVFOYLTRELLQSLASGNDCTTSTACMCVGL 389  
Db 307 WC-----PSEFLQTPVMSYLGA----- 326  
Qy 390 SPVLTSLEHLTGNOODIITRACBVLADPSLPELFWGTEPTSGLOIIDSVCGMFHLIS 449  
Db 327 -----HRRPPQRQVLSKFRWOMDLP-----PTIYI 354  
Qy 450 PLQLRLALVSGSKTAKKVSFLDQMSFYNELYKHKPHDVISHEDGTLRQTPLYLPL 509  
Db 355 PLYKMLQSLANGPOCAHYCFSL----- 379  
Qy 510 GGQTNLRIPQGVGVMDLDRAYLVREXSYSSWTLFTCEIEMLLHVVSTADVIGHQCRV 569  
Db 380 NGSSHVENIQAGGSP-----VSWEHFHSLLYH--EHLKRD.PSADSVQY--RH 426  
Qy 570 KPIIDLVHKVISTDLSADCLPTSRIMYLQ--RLTVVSP--PRDVYASCMVC-- 621  
Db 427 LPSRGITOK--EODGLIA--FLQLTSTIITWSENARLALCEHPQWTPVAVIIGLQCSIP 482  
Qy 622 -----LTVLAA--RNP--AKWMTDLRHTGFLPFVAPVPSLSOMISABGMAGYGN 669  
Db 483 PVLKAEELKTLAFAFGSPBIAASLWOSLETOILQTVRVPSQQAIGIEVB----- 533  
Qy 670 LLANSEPOGEGYVTIAFLRLITTLVKQSLGSGTOSQGL-----VPCVMEFLKEMLPZYHK 724  
Db 534 --LNEIESCEBEPPLTRAFQOLISTLVESFPNLAGARPPGFDDYLOGLRBSVFLRPT 592  
Qy 725 WRYNHGVREQIGCILELHAAILNLCHETDLSHSTPRL-----QFLCTCS--LAYTE 776  
Db 593 RAYRRAAEKWEVAVLEVFYKILR-----DYEPQLEDPVQFVELQSEETIAYXP 643  
Qy 777 AG-----QTVINIMGVGDPTIDVMAAOPRSDAEOGGQGLIKTKVK--A 821  
Db 644 PGFSIMYHLNESPMLELALSLLEBGVKQLD--TYAPPECKHLEKRAVQVQCLMLNLTLQK 702  
Qy 822 FSVTNVIRLRKPSNVVSPLEQAL--SQHGAHGNLLAVLAKYIYHKHPALPRLATQ 878  
Db 703 ENLPMDLRBSGLALIVSPLEBQLOGINPRTKKANVMI--ARYLYHGNN--PELAPES 759  
Qy 879 LKRLATVAPMS-----VYACIGNDAA--IRDAFLTLQSKI 913  
Db 760 AKILICISNSNIQYKMGVDPFTHDQSVSQKLAGFEICLDYDTEBFYVABGESLEKEL 819  
Qy 914 EDMR--IKVMILEFLTVAVETQPGILIEP--LNLVYVDSGDSGKESISLGMW-----SCLHAY 966

Db 820 AAINRETRHILNLLITLSERNPNLALYLLGFELKKEISTTNLODPGVLGCPRTCLHAI 879  
Qy 967 LELIDSOQOOR-----YWCPLRLHRAAIAFLHAL--WQDRBDSAMLVTRTKPFMEN 1016  
Db 880 LNIIEKTEGGDGPVAVREY--POLADLCYQVITYQLACGSDTSGPTKRYLTSODF-- 933  
Qy 1017 LTPSLFGLTSP--PSETPSPSILETCALINKIICLEIYVVKSGSLDQSLKDTLKPSI- 1072  
Db 934 --LFSQLQHLPPSNKHEHISMLSQMSWMLKTSIBELRY--SLNRQSHTRQLHL 986  
Qy 1073 -----EKRFAYWSGYKSLAVHVAETEGSSCTSLLEYOMLVANPMLIITATADIMHL 1127  
Db 987 LDDMPKVPKSPDEGGM-----EDENRSYSGFLHPTATKVRKILSI----- 1028  
Qy 1128 TDSV-----VRQFLVDYDQTKALLVPAVNCIR--LGSNCTLLILLRQKREL 1179  
Db 1029 LDSIDFSGEIEPQLQPFDDAQIBQVY--ANCEHKNLQOQTVCNVLL----- 1075  
Qy 1180 SVDEILGPLTEILLEGVLQADQQLMEKTKAYKVSAPIT-----VLQMKKVSDDIPQSOV 1235  
Db 1076 --HRYLVAVVNAALQMAAIGRPLMBEISTIIQYVGRNKLQCLHAKRBALESWRQV 1133  
Qy 1236 LNVCEILOEVIALPDQTRHSLALGATBDKDSMETDDCSSRRHDDQGVCVLGLHLAK 1295  
Db 1134 EILITACPOELIOABDR--QLIIRDLQDVHDKVLD-----EAAQ 1172  
Qy 1286 ELCEVDEDEGDSWLQVTRRLPLPLTLTLTLESLAMKQULHTEATNHLTLARQOQAT 1355  
Db 1173 EL-----MPVAAQAVFTL-----TAHLSQA-----VREOROP 1200  
Qy 1356 AVAGAGITQOSICLPILSVYOLSTNGTAQTPASR-----KSLD--APS 1396  
Db 1201 LVSGRGEQO-----YAFMLDSSLTSSPAESRPVGRASIGDSSMIHILKKLDPILKT 1253  
Qy 1397 WPGYVRLSMSIMEQLKTLRNFPLPEALDVGVOERTLQCLNAVRYOS----- 1446  
Db 1254 GGGFORVYTHLYGSLLYVLIQAORPDEBDTLEAAKTWERTLTAPEDEVPSKLORENMAII 1313  
Qy 1447 -----LACLEAD-HWYGFILQS-----NMEMHPTLPOLMEDIQNLGYL 1488  
Db 1314 ESYGAALMEVYCRACDHEIGRMALALDLRIVSVDOHQWMLYLS-----NSGL 1365  
Qy 1489 COACTSLHSRMLQHYLQNGKNGDGLPSAVQVORPSPASAAAPSSKQPAADTEASEQ 1548  
Db 1366 KVLVDSLVDDRTIQLSL-----TPQPLKALTYTESKMAFLTRVAEQ 1410  
Qy 1549 Q 1549  
Db 1411 Q 1411

RESULT 24  
Q6BLR9 PRELIMINARY; PRT; 1641 AA.  
ID Q6BLR9;  
AC Q6BLR9;  
DT 01-OCT-2004 (TREMREL. 28, Created)  
DT 01-OCT-2004 (TREMREL. 28, Last sequence update)  
DT 01-OCT-2004 (TREMREL. 28, Last annotation update)  
DE Similar to CA3071|CanuP188 Candida albicans CanuP188 Nucleoporin.  
GN ORFNames=DEHA0F122109;  
OS Debaryomyces hansenii (Yeast) (Torulasporea hansenii).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Debaryomyces.  
NCBI\_TaxID=4959;  
RN NCBI\_TaxID=4959;  
RP SEQUENCE FROM N.A.  
RC STRAIN=CB8767;  
RG GENOLABURES;  
RA Dujon B., Sherman D., Fischer G., Durrene P., Casaregola S.,  
RA Lafontaine I., de Montigny J., Marck C., Neuvéglise C., Talla E.,  
RA Goffard N., Frangeul L., Aigle M., Anichour V., Babour A., Barbe V.,  
RA Barney S., Blanchin S., Beckerich J.M., Beyne B., Bleykasten C.,  
RA Bolserne A., Boyer J., Cattolico L., Confantolero F., de Darvar A.,







DT 05-JUL-2004 (TrEMBLrel. 27, last sequence update)  
 DT 05-JUL-2004 (TrEMBLrel. 27, last annotation update)  
 DE AER284WP.  
 GN Name=AER284W;  
 OS Ashbya gossypii (yeast) (Eremothecium gossypii).  
 CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 CC Saccharomycetaceae; Saccharomycetaceae; Eremothecium.  
 NCBI\_TaxID=33169;  
 RX PubMed=15001715;  
 RA Dietrich F.S., Voegelé S., Brachat S., Lechar A., Gater K., Steiner S.,  
 Mohr C., Pohlmann R., Ueidi P., Choi S., Wing R.A., Flavler A.,  
 Rafferty T.D., Philippen P.;  
 RA "The Ashbya gossypii genome as a tool for mapping the ancient  
 RT Saccharomyces cerevisiae genome."  
 RL Science 304:304-307(2004).  
 DR EMBL; AER16897; AAS52965.1; -  
 DR ACD; AER284W; -  
 DR InterPro; IPR002129; Pyridoxal dec.  
 DR PROSITE; PS00392; DDC\_GAD\_HDC\_YDC; UNKNOWN\_1.  
 DR SEQUENCE 1501 AA; 169463 MW; ACCD26E3015829C CRC64;  
 Query Match 2.9%; Score 258.5; DB 2; Length 1501;  
 Best Local Similarity 19.8%; Pred. NO. 1.1e-07;  
 Matches 317; Conservative 220; Mismatches 561; Indels 507; Gaps 74;

332 LRLRLN-----ERTSSV-----VRKIGT-----AIOQL 357  
 174 LRLRLNVLVYVSVKAYLEETSETVDFKXVFGTTTDDLLNTPAABSLH 233  
 358 VFQYTRLLQSLASGNDCTTSTACVYGLLSFVLS-----ELHTGNODIID 409  
 234 VLDGICRLHKTLSDEG-----LYSAIMCSFLILSLNFIISLDETTHAKLVLTGMPELDIE 289  
 410 TACEVLADPSLPELWGTETPTSGLIILDSVCGMPPHLSPLQLRLALVSGSKTRAKKY 469  
 290 S---PLTSPLEKLU-----SILPAKLPLHESLIPMINL-----TSIHPEF 328  
 470 SFLDMSFYNELKXKPHDV---ISHEDGLWRROTPEKLY-----PLGQTNLRI 517  
 329 AHPEKELSTYAKKANDIDYDSDVDN--RVESTDLVYKNELVYPRPEFEDNWL 386  
 518 P--OCTGVQM---LDRAYLVWEXSYSSWTLF-----TEIEMLLHVSADVIOHC 566  
 387 PIKOTKQIIPSATGDEDVVV--FLYKSGMALLIGLILQNICETVRDSCAPPTARLQYL 445  
 567 QRVKRIIDLVHNVISTDLSIADCLPITRIMLQRLTVI----- 608  
 446 --LVSLILISSIVADTPL-----ERSTEILQHSIGVTEDDIISLIRTFEQALH 495  
 609 SPVAVIASCVNCLFVLAARNPAKWTDLRHGTFLPFVAPVSSLSQMSAEGNAGGYG 668  
 496 SRDMVLVPLCLKLTFTSMVTNPHIYKSH-----SKSDLLDRHG-KAGMT 540  
 669 NLNANSEOPQEGYGVTAFLRLITLVKQL-----GSTQGLVPCVMFLKEMLPXY 722  
 541 TLIGALELPNGEYVLTIAFLIKLADELVSLSHETAFPERMKRELLGRIVKOLLHYESY 600  
 723 HKWRNSHGVRQIQGLIELHAILANLCHETDHSHTSPSQFLCICGLATEAGQYI 782  
 601 QVMKSNINQREIQLLSTLFTKLIYSYVIGIDQSKPGDKI-----TRVLADAASFIV 654  
 783 NIMGIGVDTIDVMAAOPRSDAGEGQGLIKTKVLAFTYNNVIRLKPPSNVSP-- 840  
 655 TVF-IGTSDPD-----IRAVNSLVSV-----LVSPPT 680  
 841 LEQALSQHG-----AHGNLILAV-----LAKYIYK-----H 867  
 661 LEAIGLNGVNGAQVYVELLQKSFAPGNMLISVRLGLKLPPTLEKSIYKATDLVDAYIY 740  
 868 DPALRLAIQLLRLATVAPM-----SVYACIGN-DAAIRDAFLRLQSKIEDMRIKYM 921

DB 741 KPSLKVHTIKLFTYLVR-APWFAERPSLLALHGNHAKSLLELTFDLSPIODRYVLKN 799  
 QY 922 ILLEFLTAIV--TQGLLEFL--NLEVQSGD-----GSKFSLGMSCLHAYVELDS 972  
 DB 800 IYSPFSSLMCKQDQMSLFTLGNIVTFDDKHENKEVSK-----SILTIILK 848  
 QY 973 QOQRYWCPPLIHR---AIAFLHALMODRDSA-----MLVLRPKFWENLTPPLF 1022  
 DB 849 NALNDRPFDLGHLDLAIYAFNSWTSANRYAADPITALVRLG--AEPVNDTLK 907  
 QY 1023 GLTSPSETSR---PSILETCALIMKICLEIYVVGXSDQSLKDTLKK--PSIEKR 1075  
 DB 908 SIEQVASWASKYRLISRAICAL-----YLFTSGTTPSPFDLGRUDLATIYVP 958  
 QY 1076 FAYMSGYKSLAVHVALETGSSCSLSLEYO--MLVSMRLLIATHADIMHLDSDV 1132  
 DB 959 LFEIDGVNKDILHNLAVN-----FEKQMPGSLSTFSSTLVDRDSSPOTIYDPL 1010  
 QY 1133 RROLFLVDLDTGKALL-----LVPAVNCRLRSGMKCTLLILRLQWRELGSV 1181  
 DB 1011 MDQIF---SGDEAWFGSEHTGLAEQVVASINI---QFTYQIAAKSGALLTS- 1060  
 QY 1182 DELIGPLTEILEGVLDQDLMEKTKAKVFAPIVV---LQKEMKVSIDIPOYSQVLVN 1237  
 DB 1061 -----PIKTPVPLQDTFVDIVLHLKVNIEKVSFPFDIYLE 1100  
 QY 1238 VCETIQEIVIALPDQTRISLALGATBEDKSNEDDCCSRHRDRQDQCVGLHIAEL 1297  
 DB 1101 RIF-LCFVLYVSPKTSKQV---EKNIVE-----ITSHMKIL 1135  
 QY 1298 CEVEDGQSWLOVRR---LPILPTLITTEVLSRMKNLHFEAT----- 1340  
 DB 1136 TSSEIDFGNIASSKSAVYRPLRSILIALSL---VATGHFIEVVELLEPFRIPC 1192  
 QY 1341 --LHLLTLARTQCATAVAG---AGITOSI--CEPLSYQLSTNGTAQPSARKS 1391  
 DB 1193 KGVNLIISLISEINTTSSCGNNVITSVTXIQDILLISLF-----TXIKN 1240  
 QY 1392 LDAPSPGVYRLSNLSMEQLKTLRNFPLPALDFVGHQERTLOCLNAVTVGSLACIE 1451  
 DB 1241 LQPPK-----SFLM-----IATSLNEVDIKALMLY 1268  
 QY 1452 EADHTVGTILOSNNKEMHFLPOLMRDIOVNGIYLCQACTSLHRSKMLQHYLQNNKG 1511  
 DB 1269 SSSHL-----FRVNEPILADITLFLCEPQSVDEVAEKLTI----- 1305  
 QY 1512 DGLPSAVARVORPPSASAAPSSSKQPADTEASEQALHTV--QYGLIKLISKTLAAL- 1569  
 DB 1306 NGLPSVLE-----SPISIMIOGGGKKEPVQRHLSIWENGLSITLQLSQFG 1354  
 QY 1570 RHFTPDVCQILLDOSGLDAEYNFLFALSFTPTPDEVAAPSEFGLLATVNNAL--NMLGE 1627  
 DB 1355 AQVLECCLFVTFYSKQLSSTFSWS-----DNSLAINSIIOPTSOIIMQKMPSA 1406  
 QY 1628 LDKK-----EPLQAV-----GLSTQABSTRLLKGLMTMENCFYLLISQMR 1672  
 DB 1407 LEYQKTLNSNAKTVIDVDVIEIIPGLDTELEKEKLYRS--FT-----HLTHP--K 1456  
 QY 1673 YLRDPAVHR--DKQRMKE-----LSSELSTLLSSRYF 1706  
 DB 1457 YLNSRIVPTLEEQQLDDEBRALAFVKNTEIRKQDLSFSEP 1501

RESULT 26  
 AAS52965  
 ID AAS52965 PRELIMINARY; PRT; 1501 AA.  
 AC AAS52965;  
 DT 23-APR-2004 (TrEMBLrel. 27, Created)  
 DT 23-APR-2004 (TrEMBLrel. 27, Last sequence update)  
 DT 23-APR-2004 (TrEMBLrel. 27, Last annotation update)  
 DE AER284WP.  
 GN AER284W.

OS Ashbya gossypii (Yeast) (Eremothecium gossypii).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Eremothecium.  
 OX NCBI\_TaxId=33169;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 10895;  
 RX PubMed=15001715;  
 RA Dietrich F.S., Voegelé S., Brachat S., Lerch A., Gates K., Steiner S.,  
 Mohr C., Pohlmann R., Luedi P., Choi S., Wing R.A., Flavier A.,  
 Gaffney T.D., Philippen P.,  
 RT "The Ashbya gossypii genome as a tool for mapping the ancient  
 RT Saccharomyces cerevisiae genome."  
 RL Science 304:304-307(2004).  
 DR EMBL; AEO16897; AAS52965.1; -;  
 SO SEQUENCE 1501 AA; 169463 MW; ACCD26E3015E829C CRC64;

Query Match 2.9%; Score 258.5; DB 2; Length 1501;  
 Best Local Similarity 19.8%; Pred. No. 1.1e-07;  
 Matches 317; Conservative 220; Mismatches 561; Indels 507; Gaps 74;

QY 332 LIRHTLNP-----BETSSV-----VRKIGGT-----AIGLN 357  
 DB 174 LIRPSNPVILYWSVTLVKAYLLESTBNVDFTKNVFGTTTDLINTPAARASLH 233  
 QY 358 VFQYLRLLQSLASGGNDCTTACMCYGLLSFVLTSL-----ELHTLGNODDID 409  
 DB 224 VLDGICRLHKTLSDEG---LYSALMCSPLILSLANISLNDERTHAKIVLGMFDDLE 289  
 QY 410 TACEVLADBSLPELFWGTEPTSGLLIILDSVCGMFHLSPLLIQILRALVSGKSTAKVY 469  
 DB 230 S---PLTSELEKRL-----SILRAKPLHESLIPMINT-----TSIHEPF 328  
 QY 470 SFILDKSFYNELYKHKPHDV---ISHEDGLWRBORTPKLLY-----PLGGONTNRI 517  
 DB 329 AHFEKELSTYAKQKANDIDVLSVDN---RVSELDILVAKNELLYKPEPFEDNVML 386  
 QY 518 P--QGTGQVM---LDRAYLVWEXSYSSWTLP-----TCBIEMTLHVSSTADVLTQHC 566  
 DB 387 PLAKOTKQGIIPSAQDEBDVVV-FLYKYSGMALIGRIILQNICETYRDCAPTAALROYL 445  
 QY 567 QVRKRIILVHKVISTDSLADCLLPITRIMLQRLTIVY----- 608  
 DB 446 --LVSLIDLSSIVADTLP-----ERSTEILLQHSIGYTEDDILSLIRTEQALH 495  
 QY 609 SEPVDVIASCVCITVLAARNPAKYWTDLRHTGFLPFVAHPVSSLSOMISAEKNAGGCV 668  
 DB 436 SRDMEVLVPCLKLFISMWTNPHIYWSH-----SKSDLIDRHG-KAGWVT 540  
 QY 669 NILMNSQEPQGEYGVITIAFLRLITLVKQGL-----GSTGQGLVPCVMFLKEMLPSTY 722  
 DB 541 TLIGALELPNGEYLFITIAFIKLADELIVSESLHETAFPERMRKELGRIVQGLHIVSY 600  
 QY 723 HKMRNHSQVREQIGCLIELHALNLNCHETDLSHRTSPLOFCISLATEGQVIT 782  
 DB 601 QYWKYSNNQPREIGALLTSFTKLIYSYGGIDQSKPKDKI-----TRVLADASRIV 654  
 QY 783 NIMGIGVDTIDMVMAAPRSDGAEQGGQQLIKTVKLAFSVTNNVIRLKPSNVSP-- 840  
 DB 655 TVP--IGTDSPD-----IRAVNSLVSV-----LVSPTT 680  
 QY 841 LEQALSQHG-----AHGNNLIAY-----LAKYIYHK-----H 867  
 DB 681 LEAISLNGVNGAQVNEILLQKSFAGNMILISVRLKLPPTLLEKSTISKATDVDAIY 740  
 QY 868 DPALPRALIQLKRLATVAPM-----SYACIGN--DAALIRDAFTRLQSKIDEMRIYVM 921  
 DB 741 KPSLKVHVIKLFYLVV-APWFAERPSSLMLGHAAASLLELTFDSSPIODRVLYKN 799  
 QY 922 ILBEETVAVE--TOPGLIELFL--NLEVDQSGD-----GSKKPSLGMWSCLHVLLELDS 972  
 DB 800 IYSPFSSIMEKQKQMSILFLTGNIIVTDDSKHEKVGSK-----SILTIILKR 848

QY 973 QQQDRWCPELHR---AALAFHLNMDRPSA-----MLVLRTRKFWENTLSPLE 1022  
 DB 849 NALNMDRPEPDLGSHLLDIALISTAFNSWTSARYADKDTALVRLK-ALFEPNVNTLK 907  
 QY 1023 GTLSPPSESTSE-----PSILETCALIMKIICTEIIYVVKSGSLDQSLKDTLKK---FSIERK 1075  
 DB 908 SIEQVYASMSKRYLSRIALICAL-----YLFSTGTSPIFDLLGRDIALTVNP 958  
 QY 1076 FAYWGYVKSALVAHVALEFGSCTSLBYQ---MLVSANRMLLIATTAHMLTDSVV 1132  
 DB 959 IPELDGYNKDLATNLAIVN-----FEKOWGSLSTFSTLSTLVDNSFOTIITYDLEL 1010  
 QY 1133 RQGLFDVLDGTKALL-----LVPSVNCILRGSMKCTLLILLRQMKELGSV 1181  
 DB 1011 MDQYF-----SGDEANFGSESHGLRQVVASINL---QVTVQISANSGALLTS- 1060  
 QY 1182 DEILGPLELLEIGVLQADQOLMEKTKAFVSAFITV---LQMKMKVSDIPQVSOLVYN 1237  
 DB 1061 -----FLKKTVPVLPQTFVDIVLHLKVNIEKVSSEPFVDIYLE 1100  
 QY 1238 VGETQAEVIALFDOTRSLALGATBEDKDSMETDDCSRRHRDQDVCVGLHLAKEL 1297  
 DB 1101 RIE-LCFYLVYSFMKTSKNVT-----EKNLVK-----ITSHMKIL 1135  
 QY 1298 CEVDEDDGSMLOVERR---LPILPTLLTLEVSLSMKONLHTEAT----- 1340  
 DB 1136 TSSEIDPLGNIASSKSKAYRPLRSILIALSL---VATGTHFELVVELLEPFERJPC 1192  
 QY 1341 --LHLLTLARTQCATNAVAG---AGITQSI--CLPRLSVYOLSTNGTAQTPSASRKS 1391  
 DB 1193 KGVNLIIEILSEINNTSSCCANNVSYTSVTKIDILLALLSLF-----TKIKN 1240  
 QY 1392 LDAPSPGVNVLMSLMEQGLKTLRYNPLBALDFGVGHQSRITLOCLNAVPTQSLACLE 1451  
 DB 1241 LQPKR-----SEKLM-----IATSLAEVDTIKALNLY 1268  
 QY 1452 BADHTVGIQLSNFMKEMPHLPLQMRDIOVNLGYLCOACTSLHSRKLQHYLQNRNG 1511  
 DB 1269 SSSHL-----FRVNEPPIADITILFLCEPSSVDBVAEKLI----- 1305  
 QY 1512 DGLPSANQVRQRPASAAAPSSSKOPADTBASBOQALHTV-OYGLKLSTLTAAL- 1569  
 DB 1306 NGLPSVLE-----SPISIMIQGGIKRQVPRILMSINGLSIILQLSQFG 1354  
 QY 1570 RHFPDVQQLIDQSLDAEYFLFALSPFTPTDSVAPRFGTLATVNVANL--NMIGE 1627  
 DB 1335 AQVLPECLFTYYSKQLSSTIFSMS-----DNSIALSNSIITQDSQIIMQKFFSA 1406  
 QY 1628 LDKKK-----EPLTOAV-----GLSTQABGTRTLKSLMFTWENCFYLLISQMR 1672  
 DB 1407 LEQKVLNSNAKTKVIDDVVIEIIPGLDTLEKRELKLYNS--FT-----HLLTHP--K 1456  
 QY 1673 YLRDPAYHPR-DKQRMQOE-----LSSELSTLSLSIRYF 1706  
 DB 1457 YLNSRIVPTTLEBQRLQDEBTRAFAVKNTEIRKLODSLSPSEF 1501

RESULT 27  
 Q6FP54 ID Q6FP54 PRELIMINARY; PRT; 1670 AA.  
 AC Q6FP54;  
 DT 05-JUL-2004 (Tremblrel. 27, Created)  
 DT 05-JUL-2004 (Tremblrel. 27, Last sequence update)  
 DT 05-JUL-2004 (Tremblrel. 27, Last annotation update)  
 DE Similar to ep|P52593 Saccharomyces cerevisiae YML103c NUP188 nuclear  
 DE pore protein.  
 GN ORFNames=CA6J0J0634g;  
 OS Candida glabrata (Yeast) (Torulopsis glabrata).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; mitosporic Saccharomycetales; Candida.  
 OX NCBI\_TaxId=5478;  
 RN [1]  
 RP SEQUENCE FROM N.A.



Db 1539 SNOIGANL-----SMPDMSMSSTLMTVETKIIIL-----LOKRL 1574  
 Qy 1633 EPILOVANGLESTO-----AEGRTIKSLIMPTMENCPLLISQAMRYTPDAVHR-----1682  
 Db 1575 EALINKVNSVTSRNNSTQNPNDVKQIPGLVSEBODYKQINSALKKLL-----THPKYLSNR 1630  
 Qy 1683 -----DKQRMKQELS--SELSSTLLSS 1701  
 Db 1631 IVASTTDECHMLEIEISRNELSTIIISN 1657  
 RESULT 28  
 ID 184 SCHPO STANDARD; PRT; 1564 AA.  
 AC 09P7M8; 09Y8G4;  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 05-OCT-2001 (Rel. 40, Last sequence update)  
 DT 05-OCT-2004 (Rel. 44, Last annotation update)  
 DE Nucleoporin nup184 (Nuclear pore protein nup184).  
 GN Name=nup184; ORFNames=SPAP27G11.10c;  
 OS Schizosaccharomyces pombe (Fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 OC Schizosaccharomycetaceae; Schizosaccharomycetaceae;  
 OC Schizosaccharomycetes.  
 OK NCBI\_TaxID=4896;  
 (1)  
 RP SEQUENCE FROM N.A., FUNCTION, AND SUBCELLULAR LOCATION.  
 RC STRAIN=972;  
 RX MEDLINE=93318821, PubMed=10388805;  
 RA Whalen W.A., Yoon J.H., Shen R., Dhar R.,  
 RT "Regulation of mRNA export by nutritional status in fission yeast.",  
 RL Genetics 152:827-838(1999).  
 (2)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RX MEDLINE=21848401, PubMed=11859360, DOI=10.1038/nature724;  
 RA Wood V., Gilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
 RA Sgourou J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
 RA Brookes K., Brown D., Brown S., Chillingworth T., Church C.M.,  
 RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,  
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
 RA Holroyd S., Hornby T., Howarth S., Huckle E.J., Hunt S., Jagsels K.,  
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinovitch E.,  
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
 RA Woodward J., Wolkeat G., Aert R., Robben J., Grymoprez B.,  
 RA Welljens I., Vansteelandt E., Rieger M., Schaefer M., Mueller-Auer S.,  
 RA Gabel C., Fuchs M., Fritz C., Holzer E., Moesl D., Hilbert H.,  
 RA Borzym K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,  
 RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,  
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Leleau V., Mottier S.,  
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Huxel S.M.,  
 RA Lucas M., Koehn M., Galliard C., Tallada V.A., Garzon A., Thode G.,  
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsberg S.L.,  
 RA Centuti L., Lowe T., McCombie M.R., Paulsen I., Potashkin J.,  
 RA Shpakovski G.V., Uesery D., Barrett B.G., Nurse P.,  
 RT "The genome sequence of Schizosaccharomyces pombe.",  
 RL Nature 415:871-880(2002).  
 CC - FUNCTION: Interacts with pom152 in the core structure of the  
 CC nuclear pore complex (NPC). Involved in the export of mRNA.  
 CC - SUBCELLULAR LOCATION: Nuclear pore complex.  
 CC - CAUTION: Ref.1 sequence differs from that shown from position 1560  
 CC onward and is longer (1628 AA) due to a frameshift.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; AF055035; AAD4830.1; ALT\_PRAIR.  
 DR EMBL; AL157917; CAB76031.1; --  
 DR GenBank; SPAP27G11.10c; --  
 KW Nuclear protein; Transport.  
 FT COMPACT 1219 1219 S -> P (in Ref. 1).  
 FT SEQUENCE 1564 AA; 176962 MW; 2483AE338C30B7 CRC64;  
 Query Match 2.6%; Score 233.5; DB 1; Length 1564;  
 Best Local Similarity 19.4%; Pred. No. 5.4e-06;  
 Matches 332; Conservative 250; Mismatches 663; Indels 469; Gaps 74;  
 Qy 34 SCIEALINKHWRRLLEG-----LSYKPP-----SPSSAEKANKOVASPLKELGRISKPL 86  
 Db 43 SPVTGTLSN--KTGEBAKLSTYDSHSISKQLESVK-----KISDIT 86  
 Qy 87 GLDERQSVQL-----QCTYQEDYGRTRDSVKTVLDERQSVQLIKIADYYEERTC--IL 141  
 Db 87 GYNEAQVAVYLVVHQEYELNTQYFQOLDN--DSVLAQEFQR-----YABIIISCKVYL 137  
 Qy 142 RCVLHLLTYPDERRHRYVEYADCVDLKELVKKYQ-----179  
 Db 138 AFLQACTDADSKH-----KATRLIVISIFQTAQSGBNASTPISFCVRII 186  
 Qy 180 -QFELYTEAPWETHGNLNTERQVRWFQVQCLREOSMLLEIFLYAYEEMAPSDLLV 238  
 Db 187 DYLSKMTSOAPAPASLTENG-----EPAISQWYFHHNLQQLRLVFL-----STYSLV 236  
 Qy 239 LTKMFKQGFSGRTNNHVD-----TMDPVNIRIGYPSALLIVEGMDIESLHKALD 293  
 Db 237 CNSEMAISWNCMKKTVLHDQEPMLDIDTGFSCKEITNVAIIISINFLSKQVLSF 296  
 Qy 294 RREHLPADGLICQDNCMLTFEGDIPHHAPVYLAVALHRLTL--NPE-----ETSSV 346  
 Db 297 KDNPSFPLSGNTIISLMDITQUNDSIGAAVSLTGIALHLNSPNDIPLIONSVV 356  
 Qy 347 -----RKIGTAQIQLNVQYVLTLLQSLASGNDCTTSTACMVCYGLSFVL 394  
 Db 357 SKKILNPONSFOMLIILALKYDFLTHRIISL-----EDDPYIDGYSKIMATLF 408  
 Qy 395 SLEHTGNOQDIDINCEVLADPSLELFWGTEPISGLIIDSVCGMPPHLSPLQL 454  
 Db 409 SSASVYKESDSTMLCATTFKTPQVQLF-----ENNSVTRLNFPARARPPFYSQVLL 465  
 Qy 455 LRALVSGKSTAKKYVSEFLDKMSFYBNLYK--HKPHVISHEDGT--LMPRQTPKLLYPL 509  
 Db 466 LIPTFACLTQKQVSSSLHMTTFQSLPESGFAVEIIPBPNTGNALIELQSSHLDSY 525  
 Qy 510 G-----GQNLRLPQGVQGVN--LDDRAVLVAKESYSWTLPTEIEMLHVGSTADYIQH 565  
 Db 526 GFPPPNARSPLPGTGRIVSVDTYPVVMMDLNSLMEAVGSLN--YIVNGSLNSH 582  
 Qy 566 CORVKPDIIDLVHVIYSTDISADCLPITSR-----IYMLQRLTYTISPPVVIAS 617  
 Db 583 KSVVLTVLSSSVLPQTVSGACELVHLSBGDLGELDPNTVLCDLIDYFLSLVEDAD 642  
 Qy 618 ---GVNCLTVL-----AANPAKWTDLRHNGFLPFVHAPVSSLSQMSIABGNGGYG 668  
 Db 643 YQICVSLRLRLRFTFAATD--VW-----AYVRSIYCVSGSKGIS--LE 664  
 Qy 669 NLALMNEQOGEGVGTIATPLRL-----TTLVKGQLGSTQGLVPCVMFVLEKML 719  
 Db 685 DVLFEDYESINGVYDFLAFDFYELLDNLCISVSVPDPFSIRLKTDFVBARFLCEVF 744  
 Qy 720 PSYHKRRYNSHGRBQIGCLILELTHALNLCHEIDHSSHT-----761  
 Db 745 ANLDMKVARIIQOYQIGRPSALITKLANVFGIEYFPKTVNKTPLRLSHYIQ 804  
 Qy 762 -----PSIQFLCISLAYTEAGQTVINIMIGVDTIDVMAAPRSDAGEQ 808

Db 805 RFLVQDSNRNYLHPLLSVMDLNLITDITPSTI-----SSPRAKAAK-- 846  
 Qy 809 GGGGLLTKTVLAFTVNNVI-----RLKPPSNVSPLEQALSGHAGNNLAVLAKY 862  
 Db 847 ---MWLIS---SFCAMKTLICLRGFLNLKP-----SELERELPERSDLPFN----- 886  
 Qy 863 IYHGHDPALPRL-----AIGLKLRLAVAPM-----SVVACLGDAADAIDAPLRLQS 911  
 Db 887 ---CLPRLCCICLADITQLLSAL-ILAPMSEFTPLLYMINSTDIVGVCQILTN 938  
 Qy 912 KIEDMRKIMLEPFLTVAVE--TOPGLIELFLNLEVKDSDSKERSLQ--MNSCLHAV--- 966  
 Db 939 PIQSTNIEGSAWKLSSIMKQOQGLAVLFS-----GKKPRLDRMSLNNVVO 989  
 Qy 967 ---LELIDSOQODRYWCPPLHRAIAFLHLMQDRDSAMVLTKPKFMENTIS-- 1019  
 Db 990 LTSKSLISIAEKRLDPSINDILSQVPEEFPLSRNFWTSLGMLQOENAFMNRIVDAI 1049  
 Qy 1020 ---PLFGTLPSPSESEPSILETCALIMKICLEIYY-----VYKSLQSLKDT 1066  
 Db 1050 KLPLTVKLDGLSSVAAQDLYLAAHATRTIYALQLHMSKLKNSSSKJIIIDPLKDSMKOL 1109  
 Qy 1067 LK--KFSIEKRPAYSGVYSIA-----VHVAE----- 1092  
 Db 1110 VQHAFTIT--AYDSNHNALTRAFKHNGDLHISDLNNTGLFLRYGDNFYNNIKLAN 1166  
 Qy 1093 ---TEGSSC-----TSLELYQ--MLVSAVRMLLIITTHADIMHLTDSVVRQOL 1136  
 Db 1167 MLNLTEDTSPKISMMMSANENSLDDAQAALLRSV--SIFICAFVFEKEDATLSTLEKI 1225  
 Qy 1137 F-----LDY-----LDGTQALLLVPSVNCAL--GSMKCTLLIILBOMK-- 1175  
 Db 1226 MKWVLSLAEDTIDVNVVQELSAERPAALVFRISQOTLAIPISENVEKHLQSLITLTKAI 1285  
 Qy 1176 ---REIGSVDEILG-----PLTEILEGYLQADQOLM--EKTAKYPSAFIT--VLQMKEMK 1224  
 Db 1286 TTYTFSITEDSNGEMAYTRPLHAYNTL--NRLSEKENTLSVGFVSGLDLQCRK 1342  
 Qy 1225 VSDIPQYSQVLNWCETIQOEVIYALFDQTRHSLALGATDKDSMETDCCSRHRDQD 1284  
 Db 1343 LSQI--PEKAVINPTEIYGVDIVL-----NSLHK----- 1370  
 Qy 1285 GVCVGLHLAKEL--CEVDEDEGDSWLOVTRRLPIPLTLTLEFSLRKNONHFEATLH 1342  
 Db 1371 ---CYNSHLIRGLQSLIYSINDSP--SYDNCL-----RLFSWSHSLVDGQYFADALIS 1422  
 Qy 1343 LLLTLARQOGATAVAGAGITQSIGPLLSYQLSTNGAOTPSASRKSLLAPSPGYR 1402  
 Db 1423 FLTICSSPAGBOJIVNNGFFYSIMESPLST--ALSTGGLGLDGS---SIQYKIN---IR 1474  
 Qy 1403 LMSMIEOGLTLRLRYNPLPEALDPVGHQERTLOCLINAVTVQ-----SLAC 1449  
 Db 1475 GILPFLPIVIVKFLGRINNDREPV-----LLAFPOQYVALMWCPPSSISLAS 1524  
 Qy 1450 LEEADHTVGTILQLSNPKEMHFLPLQMLRDIQV 1483  
 Db 1525 IDES---FMIVLPLDLQCFN---PALQBIRL 1551  
 RESULT 29  
 Q6FP09 PRELIMINARY; PRT; 1710 AA.  
 AC Q6FP09;  
 DT 05-JUL-2004 (Tremblrel. 27, Created)  
 DT 05-JUL-2004 (Tremblrel. 27, Last sequence update)  
 DT 05-JUL-2004 (Tremblrel. 27, Last annotation update)  
 DB Similar to sp|P47054 Saccharomyces cerevisiae YOL039c NUP192.  
 GN ORFNames=CAGL0J00759g;  
 OS Candida glabrata (Yeast) ('Torulopsis glabrata').  
 CC Baktaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 CC Saccharomycetales; mitosporic Saccharomycetales; Candida.  
 OX NCBI\_TaxID=5478;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN=C86138;  
 RG GENOLAYURES;  
 RA Dujon B., Sherman D., Fischer G., Durrens P., Caaregola S.,  
 RA Lafontaine I., de Montigny J., Marck C., Neuvaglisie C., Talla B.,  
 RA Goffard N., Prangeul I., Aigle M., Anthouard V., Babour A., Barbe V.,  
 RA Barney S., Blanchin S., Beckrich J.M., Beyne B., Bleykasten C.,  
 RA Boismarie A., Boyer J., Cattolico L., Confarotieri F., de Daruvar A.,  
 RA Desorme L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,  
 RA Hantaye F., Hennequin C., Jauniaux N., Joyet P., Kachout R.,  
 RA Kerrest A., Kozzul R., Lemaire M., Lesur I., Ma L., Muller H.,  
 RA Nicaud J.M., Nikolaki M., Ozcas S., Ozler-Kalogeropoulos O.,  
 RA Pellenz S., Porter S., Richard G.F., Straub M.L., Suleau A.,  
 RA Swenne D., Tekala F., Mesolowski-Louyel M., Westhof B., Wirth B.,  
 RA Zenhou-Meyer M., Zivanovic I., Bolotin-Pukhara M., Thierry A.,  
 RA Bouchier C., Caudron B., Scarpelli C., Gallardin C., Weissbach J.,  
 RA Wincker P., Souciet J.L.,  
 RT "Genome evolution in yeasts",  
 RL Nature 430:35-44(2004).  
 DR EMBL/CR380956; CAG60692.1;  
 SQ SEQUENCE 1710 AA; 195270 MW; 4915517790EC73 CRC64;  
 Query Match 2.6%; Score 232; DB 2; Length 1710;  
 Best Local Similarity 18.1%; Pred.No. 7.7e-06;  
 Matches 325; Conservative 281; Mismatches 602; Indels 590; Gaps 84;  
 Qy 17 SREMTILGSAARELSQIAELNKRRLLEGLSTYKPPSPSAEKVANKOVASPLK 76  
 Db 21 SSELFTQL--PDKEKLKPPK--NTSRRNLKELQDXTDDGSI--FKLNQDPTISI- 73  
 Qy 77 ELGRISKFLGDEOSVOLQCYLOEDYGRTPSVKTVLQDEROS---QALTKIADY 132  
 Db 74 ---ISLDELHDELAVNL-----VLNSPSSQIYASVNVNGKISFLRQYIQLIYIS 126  
 Qy 133 ---YY-----EERTCILRCYL-----HLTYFQDERHPVREYADCYDK 168  
 Db 127 ICNADVWSPYKKPFVNAETNTTILSKVLENFDIISKLSDIKQBN---NQQLDMNMP 183  
 Qy 169 LEKELSKYRQGFELKYTEPT--METHGNLMTERRQVSRMVOCLRESMLEIIFLYA 227  
 Db 184 LITGIIHKRDPFESKEDYLLSQVYVGLSKRSLINTQISS--VDLIEHSLMETNDPFV 241  
 Qy 228 YF-----EMAPSDLLVTKMP-----KEQGF----- 248  
 Db 242 YFLPLTLNLFENLNTFSDVDFGLHKKTYNELITKQBYVKKPMKVLMIFLFLIGWC 301  
 Qy 249 ---GSRQTNHLYDETMDP--FYDRIGYFSALIL-----VEGMDIESLHK- 288  
 Db 302 KEDPQNRVTKLNFSDIDDPMTIAYBELGAILIVADLSRIDESTAVQYDFRALLER 361  
 Qy 289 ---CALDREBLHOPADGLICQMDCLMTFGDIPIHRAVPLAMALLRHTLNF-- 339  
 Db 362 HLPFPIPLQVDEVNTOKYNQMSLTERSDSTKI--NGPHE-----SFSRNTYRPPS 411  
 Qy 340 ---BETSVVRKXIGGTALQNL-----VF-----Q 360  
 Db 412 MISPNNTYSKDRFTNGLASTISIMDRLNEGAVFTSDMKEISKRENTSEMFTLEB 471  
 Qy 361 YLTRLQSLASGNDCTTSTACMCYGLISFVLSL---ELHTL--GNQODIITDACEVL 415  
 Db 472 NIHHPIQTFIS---DC-----AFLLTLKQABSDLSLGSDDLDDIATK-- 513  
 Qy 416 AD-----PSLPELFWGTEPTSGLGIT--LDSVCGMFPPLISPLQL--LQAL 458  
 Db 514 ADLERFVSVFYFSKRBYVAAFWEDKESAAYGVEWASAKN--DNLMKSCFYMLMGL 571  
 Qy 459 VSGKSTACKYSPIDKMSFYNELVHKRHVDVSHEDGTLMRQPKLYPLGGQTNLRIP 518  
 Db 572 SSGKKNALNVTYTL-----GDSNFTSMKX-----IADTIGSYSKILN 609  
 Qy 519 OGTVQWMLD-----RAYLVKWEYSYSWTLFTCEIMLHLVAVSTADVIGHQCRVKPI 573  
 Db 610 LKALQAKQGRNGBLNTTLALBGLNGBRTIVPLS--SLTLTILSVADQMBPBDKLM--- 665

```

QY 574 DLVHKYISFDLSIADCLPITSRIYMLQRLTTVISPPVDIVASCVNCLTVLA---ARNP 630
DB 666 -FSLHVDVIFELTKETPL-----VACENNTISTLPIILIESDR 703
QY 631 AKYWDLRHTGFLPVAHVPSSLSQMSISAGNAGVGNLNNSEOPQGEYVTAFLRL 690
DB 704 ASIMHKLDSLT-LFKGVALSDTSQYISA-----FSSILTL-----YSDVAGFLAL 746
QY 691 ITTLVX-----GQGST-QSGCLVPCVMVVKEMLPSTHKM---RXN 728
DB 747 IRNLQIESFDCNKEFLVFGHLYPNQGLPKYKSGIMPTLEFVNMKEVFNNSDKITDSFK 806
QY 729 SHGVEQIGCLILILHAILNLCHETDLSHSTPSQLFCIGSLATBAQGVYINIMGIG 788
DB 807 RHELDGR---VLEIT-----LTCNLSFDYS---VLNLSISVG 837
QY 789 VDTIDNV---MAAPRSDGAGQ---GQGOLLITVYKLAESVTNNVIRLKPNSNV--- 837
DB 838 ANLDMLVHGENFASVQDNAAAPAMFNFLFOENVYTLGYSIASVGIDNSLKHQLOTL 897
QY 838 ---VSPLEQALSCHAHGNNILAVLAKT-----YHKM----- 867
DB 838 LRLSLSVIRERTDYERTYEEELTTIKKYQSGEFYKPKYTGHLGVSFTQVFFNLDPVT 957
QY 868 -----DPALFRLAIOLKRLA-----TVAPMSYVACIGN-----DAAAIRPA 904
DB 958 CGLVYIGLNDISIPISIMILKRLALPSTEDPPL---LGNKLTLPSTIDESARIKHA 1014
QY 905 FLTRLOSKE---DKRIKWIIEFLTVAVE---TQGLIEFLPLNVEVDGSGSKPSIG 958
DB 1015 FIDQDCPIENSNDMLKVEILFELKDSIDISSKSVHLLGFGY---DNTLTFGN 1070
QY 959 M-----WSCHAVLELDS---QOQDRY---NCPFLHRAIAPLHLMORRSMALV 1006
DB 1071 LSTFQSGHSLFHASVIRLISALNELSXYKIDPAPRLAALSQLEIMLCKSQLTSDVLT 1130
QY 1007 LTKPKFMENTLSP--LFGTLPSPSTSEPSI-----LETGALIMKICLLE-----I 1051
DB 1131 -----EHITNNEIKIKMOVDPQTSNHTLMNGINLESTQSTKARFIQSSAGAYLM 1182
QY 1052 YVYVAGSLDQSLKTLKFSIE---KRAYMSGYKSLAVHVAETEGSSCTSLLEYQMLV 1108
DB 1183 FLRYTSMIQFLNANVHTISLNGTISRAPYYSILNISISYA----- 1225
QY 1109 SAMRMLLIATTHADIMHLDTSV---RRQLFLDV-LDGTKALLVPAVNSCLRLGSMKCT 1165
DB 1226 ---RIFSLRLPLHONIT-LKEXIVBENDQLKDLFVDLDKQLRKSCGNITDFDEISSV 1281
QY 1166 LLLILLRQMKELGS---VDEILGPLTEL-----LEGVL 1196
DB 1282 LELHSKTVSVSQNSKEFMDKILSFEFVKQILTLFLVRKSFSGQALMIHNNWQVQII 1341
QY 1197 QADQOLMKTAKYVSAFITVLOMKEMKYSIDPOYSQVLNVCETLQSEVIALPQTRHS 1256
DB 1342 VSDNGLGRKR---STPI--LEVDMVTPKISVYVIDVYKABELVSLMWFLEFIYQOS 1395
QY 1257 LALGSATEDKDS---METDDCSRSRH 1279
DB 1396 NASSSGKFLDSKLANLFOVCIRGSSPLSSIKLRSDDYVYANNYLLVTLKNSGMAVIV 1455
QY 1280 RDOR---DGYC-----VLGLHAKELCEV-DEDDGSMLOVT----- 1311
DB 1456 KDLRLNBRLEVAICNDALYGEBAIKITGLLLDLSVQLGNHNGNFIWMLTMSKLFH 1515
QY 1312 ---RLPLPILITLLEVSILARKONH---FTBATLHLLTLARFOGATVAVAGITQSI 1366
DB 1516 IIRSLKATDLSLNSSTENVNENFLYELTAKATAYFLIRIARTEBGAVALIESKVPDIL 1575
QY 1367 C-----LPLLSVYQLSTNGTAQTPSASRSL--DAPSWPGVYVLSLMBQLL 1412
DB 1576 AELNPLKIDPDLGLVLPDETSSNITS---SLLRVNTTDEPLIIGKEATSVSIFELIT 1630

```

```

RESULT 30
Q7SY48 ID Q7SY48 PRELIMINARY; PRT; 2159 AA.
AC Q7SY48;
DT 01-OCT-2003 (TREMblrel. 25, Created)
DT 01-OCT-2003 (TREMblrel. 25, Last sequence update)
DT 01-OCT-2003 (TREMblrel. 25, Last annotation update)
DE Hypothetical protein zgc:63510.
GN Name=zgc:63510;
OS Brachydanio rerio (zebrafish) (Danio rerio).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
OC Cyprinidae; Danio.
OX NCBI_Taxid=7955;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AB; TISSUE=whole body;
RX MEDLINE=22388257; PubMed=12477932.
RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,
RA Diachenko L., Marsina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.R.,
RA Brownstein M.J., Uedtin T.B., Toshituki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Adamson R.D., Mullaby S.J.,
RA Boeck S.A., McEwen P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny K.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Heltón B., Kettelman M., Madan A.C., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shvachenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butcherfield Y.S.,
RA Krzywnicki M.I., Skalka U., Smalhus D.B., Schnerch A., Schein J.E.,
RA Jones S.J., Maira M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RL and mouse cDNA sequences."
RN Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=AB; TISSUE=whole body;
RA Strausberg R.;
RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC055128; AAHS5128.1; -
KW Hypothetical protein.
SQ
SEQUENCE 2159 AA; 242048 MW; FOODBBA1597E70B CRC64;

Query Match 2.4%; Score 218; DB 2; Length 2159;
Best Local Similarity 18.2%; Pred. No. 9.6e-05;
Matches 406; Conservative 300; Mismatches 653; Indels 876; Gaps 106;

QY 1 MIRSKITSVL-----SFCRSRLIMTILGSLALRELSQLIBALINKMRRLLEGSLY 53
DB 20 LLAGKEVAVSVLPFKDAASMDRSTFH---ALCGTGLIELMIGDAIAPSEFQETLFSQSL 75
QY 54 YKPPSSAASKRANKOVAPLKEGLRISKPLGLDBEGSVQLCTGLQBYRYRTSRYSK 113
DB 76 -----TLRSYVSKEVN---KKLDKSLPL----- 98
QY 114 TVLQDERQSQLILIKIADYYEERTCILRCV-----LHLLTYFQDERHRYVAYACVD 167
DB 99 -----TLSPYPLKPR-----LKCIEMLARFHLHYNQDLSLACVLPYHE--- 140
QY 168 KLEKEVSKTRQOPFELIYKTEAPTWETH-----GNLMTROVSRNVF--- 209
DB 141 -----TKVFAVVIQLFKIEDPTHKMWHGLOKPGVPLARGTLLTHCYKDLGFMDFVC 193
QY 210 -----QCLRSQNL-----LEIIFLYAYFEMABSDLL-----VLTGM--FKEGGFG 249
DB 194 SMVTNSVKAYSBELTRDGNCPQLRVITSPFASSTIVSALDAVEKTIINSTIALGLPFOVGLK 253
QY 250 SROTNRHLVDETMDPFVDRIGVFSALILVGMIDISLHKCALDRRLHQPADGLICOD 309

```

```

Db 254 S-----NLSDYTAATYM-IVCOMAKVVMBAQVDSL---SVQSRSGRTPO--LIREG 302
Qy 310 MDCML-----TRGDIPIHAPVLLAMLLHTNPBSTSVARKIGTAQL 356
Db 303 LSCIITILLONKQKVGIGKTG-----YLCAVPTLSTLOSISTVHIDPLSYLPH 355
Qy 357 NVFGYLRLLQSLASGGNDCTTSTACWCVCYGLSFLVLSLEHTLGNODIIDPCEVLA 416
Db 356 LIHVMQNDQOQNEGLSDST-----GLLOSFPQNLQ---SSNIENTAKLL 401
Qy 417 DPLSPEFWGTE-PTSGG----- 434
Db 402 E---EYVACGNELPSDDGISALNQRIOPTVRLPESRYPCALDMALENHYKNSDNEKLL 458
Qy 435 ---IILDSVCGMFPHL---LSPILLQRLALVSGSTAKKYSPIDKXSPNELY--KHK 485
Db 459 HQFISLTLSCKGYOILPESSETSLMISLHPSPVSNMA-----VDYKEILINSBN 509
Qy 486 PHDIVSHEDGTLMR--RQTPKL---LYPLGGQNTLRLPGQTVGQV----- 525
Db 510 SFDEAFPLKDLLEKIKDPSPEVLSALKALQHMGMDVEDTVSSLSILHRIKXSADWC 569
Qy 526 -----MLDD-----RAYLVMEYSYSMTLF-----TCEIEMLAHV 556
Db 570 PVLKEAVRVLDDPRIIEGNPDLKAYI-----SMELLPLVMTRAPBECELOMTSAI 621
Qy 557 VSTADVIQHCORVKEPID--LVHKVISTDLSIADCLPITSRIYMLQRLTTVSPVVD 613
Db 622 TETLISQH---PLTQMAKVLKAVLAKTSESD-LGAVAN-----EMLTTLV----- 664
Qy 614 VIASCNCLTVLAARNPAKWMTDL-----RHTGFAPFVAPVSSISQWISAEQMA 664
Db 665 -IKNDAN-MDAATKANTLENVCDILSRQSSVDRBAAVVPSALQISQSTESQHLHT 722
Qy 665 G-----GYGNLL-----MNSBQPGEXGV 683
Db 723 AQSVYKLEPILLQAVTIOPEQVSDQPADBCLPVCVALGEFQIKISGLSABQBG--L 780
Qy 684 TIATRL-ITTL-----VKGO-----LGSTQSG-LVPC 710
Db 781 LLSLRLITTLKCPDSTFKGEPWNNPEKMETTCCYRLLCRLDVAVISGASQGPLAPC 840
Qy 711 VMPVLKEMLPKY-----HKMRYNSHGVRQIGCLLEIHA----- 746
Db 841 FRSLQPLQVHNPVLFKFLSLSWGNSN-LGQUDCRVSAILOTOALYVGRAPLSS 899
Qy 747 ---ILNCHETDLSHSTPSLOFLCISLAYTEAGQVINIM----- 785
Db 900 QPVKTLNLLASDS--SPVVPSL-LVCVCS-GVCEVRBAALAVLQGLSGLVSPHYPIVEK 955
Qy 786 ---GIGVDTIDVMA-----AOPRSDGABQGGQGLIKTVKCAFSTNNVIRLK 832
Db 956 LLKSEBIIADBSYLTQALSKFEBAVSRKDKKLASVEQLQCLQSGPFC----- 1006
Qy 833 PRSNVSPLEQALSGHAGNNLIYVLAKYTHKDRPLRLAIDQLKRLATVAPMSVYA 892
Db 1007 -PSTYSKTLRLQD--VHGEVLSVLL-----PAVERLLBOCAPDSCTFLP----- 1050
Qy 893 CLGNDAAIRDAFLTRLOSKEIDMRIKWILEFLVAVETQGLIELFLNLEVKDSDGS 952
Db 1051 ---DEALL---LQLLSKSEMSAPLIVD-----PRCLEVPIR----- 1083
Qy 953 KEPSIGWMSCLHVALELIDSGQODRYWCPRLIHRAIAF-----LHALMDRRDSAML 1005
Db 1084 -----ALHTSARPYPIPSQITALEQ----- 1105
Qy 1006 VLKTRP-----KFWENTLTPLF-----GTSPSBSEBPIELCALIMKICLEY 1053
Db 1106 -ITPFPFALGDEKIQOKILISILPDLVGNKSPACQNSINVPKTIADCELVANELP 1163
Qy 1054 V---VKGSLDQSLKOTLKPKFSEIKRFAVWGSVXSIAVHVAETEGSSCTSLLEYQMLVS 1109

```

```

Db 1164 ADKQVATATVQOTRBSMRK-----TQDTSGA-----VPEBSVVS 1198
Qy 1110 AMRMLLIATTHADIMELTDSVVARQFLVDLGTAKLLIVPASVNCILRSGMKCTLLI 1169
Db 1199 WPRVTLLI-----BLQHKKKLRAQYLIVRLPYLRLRCLEBPAAABENIETTQLLIC 1253
Qy 1170 LLRWKRELSVDEILQPLTEILEGLVQADQOLMEKTKAKVYSAFIVYLQMKKQVSDIP 1229
Db 1254 LNAVQCK--LSBEGGPIS--KDVLEBDKFME-----LVVQCVRVSBMP 1293
Qy 1230 Q-----YSQVLYNVCETLOBEVIA 1248
Db 1294 QTHRRALLLGLAGLPEPEKYLNIIMPITFMGANIMELDDTYSFOVIN--KTQVAVIPA 1351
Qy 1249 LPDQTRSLASGATBDKDSMETDSCSRSHRDQDVCVGLHAKELCEVDEBDGSM 1308
Db 1352 LI--KAHE--GSGSGBGNET-----VVAQIINVPDALPHV 1385
Qy 1309 QVTRRLPLPLTLTLEVS-----LRMKONLHFEATLHLTLARFQOQATAVAGA 1360
Db 1386 PEHRLPLLSQMLSTLGRSRLVLMLELFRQ--HVQ-----TSAGATGAKE 1432
Qy 1361 GITOS-----TC-----LPLSVOLSTNGTAQTP-----SASRK--- 1390
Db 1433 AVERRDDPWLIVCCBEVKEQULTSLIKILOYIMTLPODBEABEKKKPRGRSAVKDET 1492
Qy 1391 -----SLDAPSWGVYR--LSMSLMEQLKTLRYNPLPALDFVGVOERTLOCL--- 1438
Db 1493 VSDILFVETISGDLNHFKISISFMAQLAS--DGFVGKADCELD--TESTILOALQOD 1549
Qy 1439 ---NAVTVQSLA-CLEB-AD-----HTVGFILQLSNFMKWHFHP 1475
Db 1550 LLEVLRATYQVAVACVEDNADKPTAKFWBALLSKSYDTLDVYNALLPMDTFI----- 1601
Qy 1476 QMEDIQVNLGYLOACGSLSHSRKLOHTYQNRNGCLPBAVAVORVGRPPSASAPSS 1535
Db 1602 TVMRGLVGN-----QLASVRKAMELLNNK-----LQOR----- 1630
Qy 1536 SKOPADTEASEQOALHTVOYGLIKILSKTLAAL---RHFTPVQCILLDOSLDLAEN 1591
Db 1631 ---TKMLKEQIT-----ALHELIGTILSYGRSHROVTADE---BEELAINQOT 1673
Qy 1592 PLFALSTPTPDESVARFGLTLATVNALMGEIDKKEPILTOAVGSTQAEGRFTL 1651
Db 1674 ALYGLKLCRNFQGDHKEF--VPVANKAVELVADKDEKXWNGSA--LLCVAEVITSTL 1728
Qy 1652 KSLM-----FTMENCFTLL--ISQAM-----YLDPAVH 1680
Db 1729 KALAIPOHLRMPAVLDTLKERKDLNNBIYLLSAVTALQASFTLPHFISPYLLDTLLQ 1788
Qy 1681 PRDQRMKQELSS--ELSTLSLSRYFRGAPSPATGVLPSPQKSTSLSKASPEQOE 1738
Db 1789 VTRLTLKRLRLTSCPOSVRLASLSSTL--ATKLPRVVLPIITKYCQSNVDAQORNL 1845
Qy 1739 PLIQLOVAFVRHMOR 1753
Db 1846 PLMNILKEHIGNOK 1860

RESULT 31
Q6MYH8
ID 06MYH8 PRELIMINARY; PRT; 1818 AA.
AC 06MYH8;
DT 05-JUL-2004 (TREMBLrel. 27, Created)
DT 05-JUL-2004 (TREMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMBLrel. 27, Last annotation update)
DE Hypothetical protein.
GN ORFNames-AFA683.010;
OS Aspergillus fumigatus (Sartorya fumigata).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiota; Trichocomaceae; mitosporic trichocomaceae; Aspergillius.
OX NCBI_Taxid=5085;
RN (1)

```



RP SEQUENCE FROM N.A.  
 RX PubMed=4998527;  
 RA Pain A., Woodward J., Quail M.A., Anderson M.J., Clark R., Collins M.,  
 RA Foster N., Fraser A., Harris D., Larke N., Murphy L., Humphray S.,  
 RA O'Neill S., Petter M., Price C., Radhikowitsch E., Rajadream M.A.,  
 RA Salberg S., Saunders D., Seegar K., Sharp S., Warren T.,  
 RA Denning D.W., Barrett B., Hall N.,  
 RA "Insight into the genome of *Aspergillus fumigatus*: analysis of a 922  
 kb region encompassing the nitrate assimilation gene cluster.",  
 RL Fungal Genet. Biol. 41:443-453(2004).  
 DR EMBL:BX649606; CAF32025.1; .  
 KM Hypothetical protein.  
 SQ SEQUENCE 1818 AA; 202590 MW; 64FE189BFCB2974 CRC64;  
 Query Match 2.4%; Score 217.5; DB 2; Length 1818;  
 Best Local Similarity 17.2%; Pred. No. 86-05;  
 Matches 315; Conservative 301; Mismatches 699; Indels 515; Gaps 74;  
 58 SPSSAKVYKANDVAS-----PKEL---GLRISKPLGLDE----- 90  
 67 SPSSAKVYKANDVAS-----PKEL---GLRISKPLGLDE----- 126  
 91 EGVVQLQCYLOEDRGTG---DSKTVLDEROSQAL-----ILKI 129  
 127 RPAALIAFPAEESTSTLSQSAAGIDVLRASLAGPALSEVLRQKAGDSDPLSENNRLRL 186  
 130 ADYYEERTCICRVLHLLTY-----FODERHYRYEVADCVKLEKELVSKYRQFREL 184  
 187 RNLVLSSESHLTKARKLALSLHNSQNDNAPRAVHGRIOQLKCAAFQOE----- 241  
 185 YKTEAPTW-----ETHGNLM-----TERQVSRWFVQCLREQSMLEI 221  
 242 -KSESESRMSFYEACIKAIKIDRLSALBEGGWLGVABSEAVENNRITLISEVLAHYQM 300  
 222 IFL-YTAYPEKAPSDLVV--LTRMEKQSGSRQTNRLVDETMDFVDRIQYFSAIILV 278  
 301 LFLQASAEIPTAELVSWLRVMDYSLSEVQVSHPCNPLRLPL-LQAFVALTYL 359  
 279 EGMDDI-----ESLHKCALDRRELHOFADGLICODMCLMPLFG-DIPIHAPVLAVAL 332  
 360 AFMKPLAMPSTIINKSTPTSPSGSPFLSKGELG-LNEIFVTAIGESTKTNPAAPSMGL 418  
 333 LRLHTLNP-----EETSSVVRKI-----GGTAIQLVNFOYVLTLLQ----- 367  
 419 LHLTRRELALSDKEIRELEQFHSAYDSFQSNTPHNSGQASLSYBELLECARPTCTA 478  
 368 --SLASGNDCTTSFACMCVGLSLFVLTSLLEHLTLANOQ----- 405  
 479 EDSIALTLTSDAMKOTMFEITVALATNVGSTSANDILTERWTRVLLDLIRVRIYVEYS 538  
 406 -DIIDTACEVLADP-----SLPELFWGTEPT---SGLGILDSVCGM---FPHLSPL 451  
 539 PEIVGSVALILEGPTKLLCSQSLSVATDPKRWIMNDPLMDELPHILAKSRFFETVPF 598  
 452 LQLLRALVS-----GKSTAKVYSFLDKMSFYNEL--YKAKPHDIVISHDG----- 495  
 599 LKLCALISKDMVNEGIPQILSEWENMDFTQIYPLFEQYETIREDENNALVTLQPL 658  
 496 TLMKRGTKLY-----PLGGQTNLRIPQGVQVMDLDRALVLRKMSYSSWTLFCE 549  
 659 AIFPSTARKQIMDHPNALVVTSSHPSTGLGVISEAKPAVIMWHEYSCLSYLSGY 718  
 550 IEMLVAVSTADVIOHCQVRKPIIDLVHKVISTDLSIDCLLP----- 592  
 719 LEEWNEENGSSGVEE-DNVAEIIGLM-----ADLLVAWVQPTNEASGAKRILEVAS 772  
 593 -----ITSRIYMLQRLTTVISP-----PVDVIAVCVCLTVLAARPAKYWTLDR 638  
 773 DGLARKSDIISLVPIFERILHNIIGPRAGDSKVSVAACLFIALTMVLPGRW----- 828  
 639 HTGFLPFAHVSSLSQMIASBGMNAGYGNLMM-SPQPOGEYGVTTAFLRLITTLV-- 695  
 829 -----PLARSSILGSDG--KGGVMTAIVATETVSGDYFLLCEVTLFQDVVD 876

QY 696 -----KQQLSGTSGGLVPCVM-----FV--LKEMLPSTYKMRKNSHG 731  
 DB 877 AASRAMVRKCPNNLSGKRTAASEWTSVPTTHMRGILLNFATWVEVFNINDIRNAE 936  
 QY 722 VAEQIGCLILELHAHILMCHETD-----LHSGHT-----PSLQPLCISLAYEAGQTVIN 783  
 DB 937 QLEINTTLAKPREMIYTTGTNTATLDSKRVTVSPSSATYLLDVARPGSRDLPRP 996  
 QY 784 IMGIVDITD-----NMAAOPRSDGEGGQGLIKTYKALFASVTNNVIRLKP 833  
 DB 997 ILRLIVDGLQTPATLHLRYLVLMKQVRS-----TLELTRLEEARVLEQ 1042  
 QY 834 PENVYSPLEQALSQGHANNLIALAKYIHKHPALPRLAIOULKL-----ATV 885  
 DB 1043 PLSL-----LEDQLFKASPLVTKLYSLDAN-----RLPVILLFILLSGALDSAN 1089  
 QY 886 APMSVYACIGNDAAIRDAFLTRLOSKLEDMRIKWMLEPLTVAVEYTOGGLIELFL----- 941  
 DB 1090 EPPSLVGHLAGASSCLFLDVLSDPKPLSDQKQLAVHNFSLSTVSKKQOMLAVYILNCS 1149  
 QY 942 -----NLEVKQSGDSK-----EFLGMSCLH-----AVLELIDSGQODRYWCP 982  
 DB 1150 SARQTMKQTDYEGGPTWKGTPFLQIALGMLNIBQMDLRAALSLIEFVSCAQENMPMATP 1209  
 QY 983 LLHRAAIAFHALMQDRDSAMLVIRTKPKEMENLTSPLFTLSPPSETSBSILET-CA 1041  
 DB 1210 -----ELRTPQPFSSLVN--YVKLTKIPSLPVDQIPATYIA 1245  
 QY 1042 LIMKICILEIYVYVKGSLDQSLAKDTLKKEFKPAYSQGVKSLAVANAETEGSSCTSL 1101  
 DB 1246 AVVADLCVYVHSAKMDQDRFTKTLIP-----LVFM--YSKD-AVDVSNVNSLHNL 1296  
 QY 1102 -LEYQMLVSNRMLIATTAADTMH--LTDVVRROLFLDVLDSCTA----- 1146  
 DB 1297 KRFEMRYSQCKL-----ADFRKTSIQPRSGRGDYDYIQLGDKLSYDFAWSGTKN 1348  
 QY 1147 -----LLVPAVYVNCRLSGMKCTLLILLRQKRELGASVDILGPLETLE 1193  
 DB 1349 QGRABEPRRANINISLVEAOS-----LHSMKFP-----ATEHCT 1384  
 QY 1194 GVLQADQQLMEKTYAKYPSAFITVLOMKEMKVS DIPQ-----YSQVLNVCE 1240  
 DB 1385 DFM-ADRE-VQKSNASVVGSL-----EANKVGPQEAIGERIQGTVEFAQML--Q 1433  
 QY 1241 TLQF-----EVALPQDTRHSLALGSAETEDKSMETDCCSRKRRDQDVCYLGHTA 1294  
 DB 1434 RLVERAGKAGAEVFGILRVWMDALRTRAYE-EALVNDDAEYR-----SVLVNVL 1484  
 QY 1295 KELCEVEDGDSMWQVTRRLPIPLTLTTEVSLRMKONLHFTENTLHLTLARTOGA 1354  
 DB 1485 ALQCHDSNPRFTTDDTSRIAMEIPSDTLVLEIKTVVAHGFSLTYLHQDPKCEKD 1544  
 QY 1355 TAVAGAGITOSICPLLS-----VYQSTNGTAQTPSASRKSADAPSPGVYRLSMGLM 1408  
 DB 1545 FALTLAIQTSLOKKNADRLYEHIIVHIBDNNTA-----RHANSLSWMAQOLAVSGPV 1598  
 QY 1409 E---QLKTRAYNLPALDPVGVGHORRTLOCLNAVTVGSL----- 1447  
 DB 1599 YGSLSTLFLVYDGLTLPMLAHLAV--EAVLTRLSCTGLTNILQPRGFPDAPRLTYI 1656  
 QY 1448 -----ACEBAHTGVFILOLSNFMKEWHFHLPOLMRDIOVNLG-----YLOACQSL 1495  
 DB 1657 WTAGFLVCLNLYHTLPTABEVAALNQBEGQLKRALESFVADRSSPSRRICLSMAE 1716  
 QY 1496 LHSRKMLOHYLQNKNGDGLPSA--VAQRVQ 1523  
 DB 1717 AYSLALIS-PLINRFREAGPSAGVDAQSIQ 1745  
 RESULT 32  
 CAF32025  
 ID CAF32025 PRELIMINARY; PRT; 1818 AA.



AC CAE32025;  
 DT 13-MAY-2004 (Tremblrel. 27, Created)  
 DT 13-MAY-2004 (Tremblrel. 27, Last sequence update)  
 DT 13-MAY-2004 (Tremblrel. 27, Last annotation update)  
 DE Hypothetical protein.  
 GN AFA63.010.  
 OS Aspergillus fumigatus (Sartorya fumigata).  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
 CC Eurotiiales; Trichocomaceae; mitospore Trichocomaceae; Aspergillus.  
 CC NCI\_TaxID=5085;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Pain A., Woodard J., Quail M.A., Anderson M.J., Clark R., Collins M.,  
 RA Fooker N., Fraser A., Harris D., Lark N., Murphy L., Humphrey S.,  
 RA O'Neill S., Petter M., Price C., Rabinowitch B., Rajandream M.A.,  
 RA Salzberg S., Saunders D., Seeger K., Sharp S., Warren T.,  
 RA Denning D.W., Barrett B., Hall N.,  
 RT "Insight into the genome of Aspergillus fumigatus: analysis of a 922  
 kb region encompassing the nitrate assimilation gene cluster."  
 RL Fungal Genet. Biol. 41:443-453(2004).  
 DR EMBL; BX649606; CAF32025.1; --  
 KW Hypothetical protein.  
 SQ SEQUENCE 1818 AA, 202590 MM, 64FE189BFCB2974 CRC64;  
 Query Match 2.4%; Score 217.5; DB 2; Length 1818;  
 Best Local Similarity 17.2%; Pred. No. 8e-05;  
 Matches 315; Conservative 301; Mismatches 699; Indels 515; Gaps 74;  
 QY 58 SPSSAKYKANKDVAS-----PKEL-----GLRISKFLGDE----- 90  
 DB 67 SPSSATKADPESKTAIAHAETTAQASDYDAKEIKADTLMSCAGIDEITATLITLMD 126  
 QY 91 EOSVQLLOCYLOEDYRGR-----DSVKTLODEROSQAL-----ILKI 129  
 DB 127 RPAARLLAFTEESTLSQASAGINDLRASLAGPALSEVLRQKAGDDSPLESENRRL 186  
 QY 130 ADYVEERTCLRCVHLITV-----FODERHRYVEVADCVDKLEKELVSKYRQFEL 184  
 DB 187 RNLYSESRSHLTKARKLLASLHNSSQNDNAPAPHVSGRIQSLKCAAAVFOR----- 241  
 QY 185 YKTEAPM-----ETHGNLM-----TERQVSRWFVQCLREOSMLLEI 221  
 DB 242 -KSSESESRSPFEACIKAIKDLISLBEGGMLGVAESEAVENMRTYLIIEVLHVQM 300  
 QY 222 IFL-YVAVFEMAPSLV-LTKMFEQGFSGRQTNRLVDETMDFVDRIIGYSALLIV 278  
 DB 301 LFLQASAEIETAEILVSWLWMDYSEFLSEVQVSHPCQNPRLILP-LQAFVALTTL 359  
 QY 279 EGMMD-----ESLHKCALDDREHLQFADGILIQDMQCLMTFG-DIPHAFLVLANL 332  
 DB 360 AFMKLPLAMPSTINKSTPTSPSOSPFYLSKEIGQ-LNEIFVTAGLESKTANPAAPSWG 418  
 QY 333 LRHTLNP-----EETSSVVRKI-----GTAIQLANFYQVYTRLIQ----- 367  
 DB 419 LHLTRELALSKELRELEQPHSAVDSPQNTPHNSGQASLSLYBELLECARPTCTA 478  
 QY 368 --SLASGNDCTTSPACMCVGLSLFVLTSLELATLGNQ----- 405  
 DB 479 EDSIALTLSDAKOMTFETIVALATNVGSTSAYVDILTERMTWRVLLDLIRVRIYEV 538  
 QY 406 -DIIDTAELVADP-----SLPELFWGTEPT---SGLGIIIDSVCGM---FPHLSL 451  
 DB 539 PEIVSSVLAILEGPTKLCLSSQSLSVAIDPRMIFMNDPFLMDELPHLARSPFYEV 598  
 QY 452 LQLEALAVS---GKSTAKVSPFDIKMSFYNEL---YKHKPRDVISHEHG----- 495  
 DB 559 LKLCALSLSKDMVNEEGIPOLISEMBNDTFQIYPLERQGYETREDENANLVLTQPL 658  
 QY 496 TLMRRQTPKLY-----PLGGQTNLRIPQGTVGVMDDRAYLVWREYSYSWTLFTCE 549  
 DB 659 AIFBSSTARQIMDHPNALVVTSSSHVPSSTLGGVISEAKPAVIMWYHEVSCSLVLSG 718  
 QY 550 IEMLHVSTADVIGHQCRVKRPIIDLVAKVISTDLSADCLP----- 592

DB 719 LBNMENGFSFGVEE-DNVAEIIGLM-----ADLLVAAWQPTFNBASSGAKRILBYAS 772  
 QY 593 -----ITRIYMLLORLTTVISP-----PVDVASCVNCLTVLAARPAKWTDLR 638  
 DB 773 DGLARKSDIISLVDPFERIILHNIGPRAGLDSKVESVACIRFIALTMVLPGRVW----- 828  
 QY 639 HTGFLPFAHVSSLSQWISAEQWAGGVNLMN-SQPOGEVGTATPRLITLV-- 695  
 DB 829 -----PULAKSSLSGSG--KGGVMTALVSAIVTSIGDYFFLECYKLFQDVDD 876  
 QY 696 -----KQGLSTOSQGLVPCW-----FY-LKEMLPYHKMYNSG 731  
 DB 877 AASRAMWKCENNLSCGTAAASWTSVPTHMIGILNFPRTWVEVNSINDMFNPE 936  
 QY 732 VREOIGCLILHAILMLCHTD---LHSGHT---PSIQFLICGLATEAGQTYIN 783  
 DB 937 ORLEINTLATPERMIYTYGTNDATLDSKVGVFSSSATYLLDVLRPOSADLPFNP 996  
 QY 784 IMGIGVDIID-----VWMAAPRSDDAGGGGQGLIKTYKLAFSVTNNVIRLKP 833  
 DB 997 ILRLIVDGLQTPATLHLRYLVLMKQVRS-----TLELTRLEBAARYLEQ 1042  
 QY 834 PSNVVSPLEQALSGHAGNNLIATVAKYIYHKDPALPRLAIQLKRL-----ATV 885  
 DB 1043 PTLT---LEDDPFRASPVLVTKYSLDAY-----RLPVILFEIILSGALDAN 1089  
 QY 886 APMSVYACLGNDAAIRDAFLTRLOSKEIDREIKWILEFLTVAVETOPGLIELFL--- 941  
 DB 1090 EPPSLVGHLAGSSCLFVDVLSQFDPKLPDQKQLAVHFPSTIYSKQOMLAAYILNGS 1149  
 QY 942 -----NLBYKQSDGK-----BSLGWMSCLH-----AVLELDSQGDVYMCBP 982  
 DB 1150 SARQTMKQTDYEGGTMGTFPFLQIALGMLNIEQMDLRAALISLEFVSCQENPMKTP 1209  
 QY 983 LLHRAIAFLHALMODRDSAMLVLTQPKFEMENLTSPLFTLSPPSESTSESILET-CA 1041  
 DB 1210 -----ELKTPQPFSSLVN-YVKLTKIPSLPVQDQIPATNIA 1245  
 QY 1042 LIMKIICLEIYVYVKGSLDGLKDTLKFSIEKRPAYSGYKSLAIVAETESSCTSL 1101  
 DB 1246 AVVADLCAYVLSAADMDDRTFYKTLIP-----LVFW--YSKD-AVVDVSAVNTSIHANTL 1296  
 QY 1102 -LEYQMLVSARMILLITTHADIMH--LTPSVVRQFLVDVDOCTA----- 1146  
 DB 1297 KRFEMRYSGCKL-----ADFRRTSLQPSLRDYYDIQGLDKLLSYDFAWSGTKN 1348  
 QY 1147 -----LLVVPASVNCLELGSCKTLLILRQWRELGVSDEILGPLEILE 1193  
 DB 1349 OGFAEFERAININSLVAQVS-----LHSMKFF-----ATEICT 1384  
 QY 1194 GVLQADQOMERTAKYKVSAPITVLOMKEMKVSIDIPO-----YSQVLAVNCE 1240  
 DB 1385 DFM-ADRE-VOKSMASVQSCV-----EANKIGVQCAIERIQGTVEFQALL---Q 1433  
 QY 1241 TLQF-----EYIALFDPTRSIALGATBEDKSMETPDCCRSRRDRQDQCVGLHIA 1294  
 DB 1434 RLVENGAKGAVFGILRAYVMDALRTTRATYR-EALVNDAAEYR-----SVANVFL 1484  
 QY 1295 KEICEVDGDSWLOVTRRLPLPFLTLTLEVSLEMKONLFTTEATLHLTLTARTQGA 1354  
 DB 1485 ALQCHLDSNPFPTPTSSRIARIBEDLVLEIVTVVAHGRSLITYLHQPOPKCEKXD 1544  
 QY 1355 TAVNAGAGITQSIQPLLS-----VYQLSNGTAQTPSARKSLDAPSPGVYRLSGLM 1408  
 DB 1545 FALTLAIQTSIQIKNADRLYEHIYTHIEDNNTA-----BHAMSLFEMADQLAVSGDPV 1598  
 QY 1409 E---OLTLTKYNNFLPEALDVGVGHQERTVLCNAKVTQSL----- 1447  
 DB 1599 YGELSIPLVGLSTLPMABEHLAV--EAVLTRLSCTRLNIIQORGGFPPDAVRLTYI 1656  
 QY 1448 -----ACLEADITVGFIIQLSINFMKMHPLPOLARIDIQVNLG-----YLQACTSL 1495

Db 1657 WTAGFLPLCLNLYLVTAEVAFLNQEGOLKRAESFVADRSGSPSRICLSMASE 1716  
 Qy 1496 LHSRKLQHYLQNGNGDLSA--VAQRYQ 1523  
 Db 1717 AYSIALIS-FLINRFREKPSAGVADSIQ 1745

RESULT 33  
 ID 09T014 PRELIMINARY; PRT; 561 AA.  
 AC 09T014  
 DT 01-MAY-2000 (Tremblrel. 13, Created)  
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
 DT 05-JUL-2004 (Tremblrel. 27, Last annotation update)  
 DE Hypothetical protein AT4g38760.  
 GN Name=AT4g38760;  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
 OC eurosid II; Brassicales; Brassicaceae; Arabidopsis.  
 OC NCBI\_TaxID=1702;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RA Bevan M., Murphy G., Ridley P., Hudson S., Bancroft I., Mewes H.W.,  
 RA Lemcke K., Mayer K.F.X.;  
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA EU Arabidopsis sequencing project;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Murphy G., Ridley P., Hudson S., Mewes H.W., Lemcke K., Mayer K.F.X.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL036565; CAB38610.1;  
 DR EMBL; AL161594; CAB80539.1;  
 DR PIR; T06075; T06075.  
 DR Hypothetical protein.  
 SQ SEQUENCE 561 AA; 62921 MW; DBCCS12226FFCD1C CRC64;

Query Match 2.4%; Score 215.5; DB 2; Length 561;  
 Best Local Similarity 22.2%; Pred. No. 1.7e-05;  
 Matches 132; Conservative 99; Mismatches 244; Indels 119; Gaps 24;

Qy 86 LGLDEBQSVQIQLQEDYRGTSDSVKTVLQDE-----ROSOQLTL 127  
 Db 9 NWLDEIQSYILVERMEGEY-GTDSVAQELQGFIDMSFLHNCQFVWVWESCEAFLL 67  
 Qy 128 KIADYYERTCILRCVHLHLYFQDERHPYVEVADCVKLEKELVSKYRQGFELVYT 187  
 Db 68 TLILSFYS--NVILEDYASVHTAPVPTKYVAYSHHTCI-----ICSVYAPRESSIKE 119  
 Qy 188 EAPWETHGNLTEROVSRWFVQCUREOSMLLEIFLYAYFEMAPSDLLVTKFKKSG 247  
 Db 120 EAVKILISG--LER-----RQSVLEDLSSCC-----PKMKGIS-- 153  
 Qy 248 FGSRTKRLVDETMDFVDRIGYSALIVGMDIESLHKALDRRELHQFADGR-- 305  
 Db 154 -GSYNSFKLAIVSVAQSHSACRVOQLMLITLIMEN-----LLQWMDVGPVF 201  
 Qy 306 ---IC-----QDMDCLM--ITFGDIPHNAPVLAWL--LKTTLAPERTSSVVRKIG 350  
 Db 202 RSGGCVSVIVQEDATISSINISEVNEAGPLVLAWVFLCLSSLPGKESRPLMDID 261  
 Qy 351 GTAIQLNVQ-----YTLRLQSLASGNDCTTSTACKCVYGLSFVLTSLHRTL--G 402  
 Db 262 HSYVHQAFEAASLYFLBILQSNLANDFDPIGSHREVVATFISAFASFEINDLEEG 321  
 Qy 403 NQODIIPACVLAADPELPELFWGTE-----PTSGLAGIILDSYCAMPHLLSPLOLR 456  
 Db 322 TLEILILILSKYQGEESLCCQFMDRKSFVDFGIFCLLFDES--BPPFSASAFIRLLS 378  
 Qy 457 ALVSGSKTAKVYEFLLDKMSFYNNELYKHKPHDVISH--DOTLWRQVPIKLLYLGQCTN 514

Db 379 SLSGSGWPAECVYNFNDKSVGVTLP-----DITSDSPADASQLVETSRPLH-IGLSG 432  
 Qy 515 LRIPQGVQVM--LDDRAYLVWEXSYSSWTLFTGEIMLHVSTADVIGHCQKVKPI 572  
 Db 433 LVIPSTNRGRIARIVSENVLVWMEYLSGIIVLIRLNKLYIGNNREAF-----VT 485

Qy 573 IDLVHKVISTDLSADCLLPITSRIYMLQRLTVVISPP---VDVIASCNCIT 623  
 Db 486 LELIRMYTENKAVCPISLNIHSFFVQSGSYNGKMSDVRVVDIICNSVRSIT 539

RESULT 34  
 ID N188 YEAST STANDARD; PRT; 1655 AA.  
 AC N188 YEAST  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 01-OCT-2004 (Rel. 45, Last annotation update)  
 DE Nucleoporin NUP188 (Nuclear pore protein NUP188).  
 GN Name=NUP188; OrderedLocNames=YML103C;  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 OC NCBI\_TaxID=4932;  
 RX [1]  
 RP SEQUENCE FROM N.A., AND INTERACTION WITH NIC96.  
 RA Pubmed=8682854;  
 RA Zabel U., Doye V., Tekotte H., Wepf R., Grandi P., Hurt E.C.;  
 RT "Nic96p is required for nuclear pore formation and functionally  
 RT interacts with a novel nucleoporin, Nup188p.";  
 RL J. Cell Biol. 133:1141-1152 (1996).  
 RN [2]  
 RP SEQUENCE FROM N.A., AND INTERACTIONS WITH NIC96 AND POM152.  
 RA Pubmed=8682855;  
 RA Nehraas V., Rout M.P., Maguire S., Blobel G., Wozniak R.W.;  
 RT "The yeast nucleoporin Nup188p interacts genetically and physically  
 RT with the core structures of the nuclear pore complex.";  
 RL J. Cell Biol. 133:1153-1162 (1996).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288c / AB972;  
 RX MEDLINE=97313268; Pubmed=9169872;  
 RA Bowman S., Churcher C.M., Badcock K., Brown D., Chillingworth T.,  
 RA Connor R., Dedman K., Devlin K., Gentles S., Hamlin N., Hunt S.,  
 RA Jagels K., Iye G., Moule S., Odell C., Pearson D., Rajadream M.A.,  
 RA Rice P., Skelton J., Walsh S., Whitehead S., Barrett B.G.;  
 RT "The nucleotide sequence of Saccharomyces cerevisiae chromosome  
 RT XIII.";  
 RL Nature 387:90-93 (1997).  
 RN [4]  
 RP SEQUENCE OF 1187-1205 AND 1611-1629.  
 RX MEDLINE=96095775; Pubmed=8522578;  
 RA Alchison J.D., Rout M.P., Martelli M., Blobel G., Wozniak R.W.;  
 RT "Two novel related yeast nucleoporins Nup170p and Nup157p:  
 RT complementation with the vertebrate homologue Nup155p and functional  
 RT interactions with the yeast nuclear pore-membrane protein Pom152p.";  
 RL J. Cell Biol. 131:1133-1148 (1995).  
 RN [5]  
 RP CHARACTERIZATION, AND NPC SUBUNIT LOCATION.  
 RX Pubmed=10684247;  
 RA Rout M.P., Alchison J.D., Suprapto A., Hjertaas K., Zhao Y.,  
 RA Chait B.T.;  
 RT "The yeast nuclear pore complex: composition, architecture, and  
 RT transport mechanism.";  
 RL J. Cell Biol. 148:635-651 (2000).  
 RN [6]  
 RP REVIEW.  
 RX Pubmed=12791264;  
 RA Suntharalingam M., Wente S.R.;  
 RT "Peering through the pore: nuclear pore complex structure, assembly,  
 RT and function.";  
 RL Dev. Cell 4:775-789 (2003).

CC -1- FUNCTION: Functions as a component of the nuclear pore complex  
 CC (NPC). NPC components, collectively referred to as nucleoporins  
 CC (NUPs), can play the role of both NPC structural components and of  
 CC docking or interaction partners for transiently associated nuclear  
 CC transport factors. NUP188 probably plays an important role in NPC  
 CC assembly and organization.  
 CC -1- SUBUNIT: The nuclear pore complex (NPC) constitutes the exclusive  
 CC means of nucleocytoplasmic transport. NPCs allow the passive  
 CC diffusion of ions and small molecules and the active, nuclear  
 CC transport receptor-mediated bidirectional transport of  
 CC macromolecules such as proteins, RNAe, ribonucleoproteins (RNPs),  
 CC and ribosomal subunits across the nuclear envelope. The 55-60 MDa  
 CC NPC is composed of at least 31 different subunits: ASP4, CDC31,  
 CC GLE1, GLE2, NDC1, NUP96, NSP1, NUP1, NUP2, NUP100, NUP116, NUP120,  
 CC NUP133, NUP145, NUP157, NUP159, NUP170, NUP188, NUP192, NUP242,  
 CC NUP49, NUP53, NUP57, NUP60, NUP82, NUP84, NUP85, POM152, POM34,  
 CC SEH1 and SECY. Due to its 8-fold rotational symmetry, all subunits  
 CC are present with 8 copies or multiples thereof. Interacts with  
 CC POM152 and NUP96.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic and nucleoplasmic side of  
 CC nuclear envelope (symmetric distribution); nuclear pore complex.  
 CC  
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL, X90580, CA66208.1, -  
 DR EMBL, U47107, AA88904.1, -  
 DR EMBL, X80835, CA56794.1, -  
 DR PIR, S47446, S47446.  
 DR Germonline, 142637, -  
 DR SGD, S0004571, NUP188.  
 DR GO, GO:0005643, C:nuclear pore; IDA.  
 DR GO, GO:0006999, F:nuclear pore organization and biogenesis; IMP.  
 DR InterPro, IPR010983, EF\_Hand\_Like.  
 KW Coiled coil; Direct protein sequencing; mRNA transport;  
 KW Nuclear pore complex; Nuclear protein; Protein transport;  
 KW Translocation; Transport.  
 FT DOMAIN 250 271 Leucine-zipper.  
 FT 35 62 Coiled coil (Potential).  
 FT DOMAIN 485 489 Poly-Ser.  
 FT DOMAIN 533 545 Gln-rich.  
 FT DOMAIN 533 541 Poly-Gln.  
 FT DOMAIN 1385 1391 Poly-Leu.  
 FT DOMAIN 1504 1507 Poly-Leu.  
 SO SEQUENCE 1655 AA; 189576 MW; B590051EBE60464C CRC64;  
 Query Match 2.4%; Score 214; DB 1; Length 1655;  
 Best Local Similarity 18.0%; Pred. No. 0.00012;  
 Matches 350; Conservative 221; Mismatches 695; Indels 610; Gaps 85;  
 QY 31 RELSGIEALNKMRLLEGISYKPPSSAEKVANKDVASPLK----- 76  
 DB 35 KOLAQIRQFLKANKNTLIESLNTIRQNTVSSGDHNLARSTIANLQINVDNDFEASQSD 94  
 QY 77 -----ELGLRISKPLGDEEOSY-----QLQCY- 100  
 DB 95 LSHAVEFFMSESRSLHIVYSLVNPDDILETFYSDNDRFNVVGLISIISSVIQNDI 154  
 QY 101 -----LOEDYRGTRDSVKTVLQDEROSQALILKADYYEERTCILRCVLAHL- 148  
 DB 155 ITASSLAHDYNNDDQMTFV-----SLVQLKKFS- -----LKFILQILQILNIMI 199  
 QY 149 -----TYFODERHPYREYKADCVLKEKL-----VSKYR-QQFEELYKTEAPTW 192  
 DB 200 LNTKVPVDIVNQWFLQYQWQF-VFECRNINSTDKSIDTSLQLYKQNFQDL----- 250  
 QY 193 ETHGMLMTEROVSRNFVQCLREOSMLLEIFL-----YAVFEMAPBDLLVLTMPKEQGF 248

DB 251 -----SYLSETLISR-----ISLFTITILLIGLNTSIAQFDI-QSPLYMDTEFDYNS 300  
 QY 249 G-SRQTNHVLDEMTDPEVDRIQYFS-----ALIVGMDLESALKKALDDRRRLH 298  
 DB 301 ALENDVATNIVNE-DPLFHPMIRHSWSPFIYRRALSSSFPDSDITKRAL----- 351  
 QY 299 QPAQDGLICQDMDCIMLTFGDIPIHNAPLYLMAALRHITN-----PEETSVVRKIGTAI 354  
 DB 352 -FASHDVLQGLANTLSELISLSPDYTYITVF-LSEPLNFIPIASISRPFAKIISSAP 408  
 QY 355 QLVNFQVYTRLLQSLASGANDCTSTACVCYGLISFVLSLEHTLANQODIIDTACEV 414  
 DB 409 EGFENFYL-----ND-----TFKKLSIIKAKPL 434  
 QY 415 LADSLPLPLFNGTEPTEGSLGIILDSVCGMFHLSPLQLARLVSGSTAKVYSPDK 474  
 DB 435 LNESLIPILNIALIDTEPANFELKDIC-----SFAVT 466  
 QY 475 MSPVNELYKHPHDVISHEDGT-----LMRQPKLLYPLGGQTN-----LRIPQTVG 523  
 DB 467 KSLANDLVYDLADITNSSSSSDIIVPDLIHKSDLVAPPLAENENCLISPKSTKG 526  
 QY 524 QVNL-----DDRVLVMEYSYSMTLFTCEIEMLHV-----STADVI 563  
 DB 527 KILTIKQO 586  
 QY 564 QHCQVRYKIIIVKAVISTDISIADCLPIISRIMLQRLTVVSPVDVYASCVCNLT 623  
 DB 587 QH-ELMISIIILKLVNVVDPKTSIEKS-----SELISYLSNLSDTASATNGAS 633  
 QY 624 VLA-----RNPAKVTDLRH-----TGFLEFVAHPVSSU-----SOMISAEKNAGGVG 668  
 DB 634 IIVQVIFBEISLQKDYTSIVQCCERMTALPNTLHVSYSLKNSDLDKTKGT--GLS 691  
 QY 669 NLNANS-BOPOGEYGVITIAFLRLITLVKQGLG-----STQSO-GLVPCVNFVLEKMLP 720  
 DB 692 NMILASVELSGDYFTIQLKLTQVFIRESLSLKNHISKSKSDIINKLIHAHIFE 751  
 QY 721 SYHKRRYNSHVRBIG-----CLIELHALINL-CHEITD---LHSSHTPLQFLICSL 772  
 DB 752 SYNNKRYNPFQKEFIAHLTLIFVDVADVFTINPHQDOLISSANKLLQLFLTPMD 811  
 QY 773 AYTBAGQVIVINIMIGVNTIDWMAAQPRDQAEQOGQGLIKTVKLAFTVNVIRLK 832  
 DB 812 SIDLAPNTLVNLLISPLNTTKILG-----DKIIGNLSKYMANSFELCTILIAIR 862  
 QY 833 PPSNVVSP--LEQALSOHAGGNMLIAY--LAKYIYHKNHPALPRLAIQLKRLATV--A 886  
 DB 863 GSNRLKFSNLEKLF-----INSSKLVQVYTLPSYVHPK-----VQIELSLYLEA 910  
 QY 887 PMS-----VYACLG-N-DAAAIRDAFLYTRLOSIEEMRIKWMLEFLYVAVER-QEGILEL 939  
 DB 911 PWNDDYPLFLSFLGAKSMAPLEKVELSDISSVPQDMNLRSLYIFPTLLBSKQGLSL 970  
 QY 940 PLNLEVKQSGDSGEFSIGMWSCLHAYVELIDSOQODRYWCPPLHRAIAVIALMODR 999  
 DB 971 FLTQFPASNKKINDSSI-----DK 990  
 QY 1000 RDSAMLVTRPKPFMENTITSLPFGTSLSPSETSEPSILETCALIM-----KIICLRIY 1053  
 DB 991 KSLITLVQKNSLLD-----STPEBVS-CKLEETIYVANTWNSKIFPKDPKF 1039  
 QY 1054 VYKGLSDQSLKCTLKPSIIEKRFAYWGSYVSLAVHVAETBSSCTSLLEYQMLVSNRM 1113  
 DB 1040 V--NSLAKLKDSKSLFOKEKELT-----RDETYSYIKKTKLIS--RI 1078  
 QY 1114 LLIITATADIMHLDTSVVRQLFLVDGTAKALLVPASVNCALSGSKCT-----LLLI 1169  
 DB 1079 VEIIFALC--LYNSTDSSEIINFNOBD-----LPELVHVFQIDGPNKTFHDELNLK 1129  
 QY 1170 LLRQW-KRELGSVDIEILGPLTEI-----LEGVLQNDQOLMEKTRAKV--PSAP 1215  
 DB 1130 FRKXKPSLELGSFOKI--PLASINENBNPGVDIPLDDIVLAKDRAMNBPBSKQTNFKERI 1187

```

QY 1216 T-----VLQKMKVSDIPOSQVLNVCETLQGEVIALPDQTHSLALSGATEDKQSMET 1271
DB 1188 TDSASLQVYVNEIETSKAMGALITTVK-----NSTVL----- 1222
QY 1272 DDCSRSHRDQDQVCLGLHLAKELCEVEDGDSMLQV--TRRLPILPILTLLEVLRLM 1330
DB 1223 -----NDGFDVLVHEFLKTLNIDFGSDKQMFQIYLERIEBFLYLSFKLSGL 1271
QY 1331 KQNLHTEATLHLTLTARQOGATVAGAGTOSICPLLS-----VYQSTNGTACTPS 1386
DB 1272 LKEKTEIEMNKKTFTLTKGEIDFTKXIGSLKNNFYRPLRSVTLVLEKVSQDRFTEL 1331
QY 1387 ASRKSU-----DAPSPGVYRLSMGLMEQLKTLKRNFLPRLADFGVHQBERTLCUNAVRT 1443
DB 1332 ISDQLLEFFELVSKGYVLLSLILQINK-----CSTRGSTDHTTQIVN----- 1377
QY 1444 VQSLACLEADHTVGTILQSNF-----MKEMHFLPOLMRDI----- 1481
DB 1378 -----LR--DNTQDILLTLFLFKITVNVPNSKNFVTLASSLNEVOTLKVILNLYSAR 1429
QY 1482 -----QVNLGYLQACSTSLHSRKMQLHYLQNRKGDCLPSANVAVRVPSPSAS 1530
DB 1430 LRRINDEPLLGQTLTFLPISLCSIEPIAAKL-----NSGLYSVLE----- 1471
QY 1531 AAPSSSKOPADTEASEQOLHTV--QYGLKILSKTLAALRHP-----TPVCOIL--LDQ 1583
DB 1472 -SPLSVAIQGGDIKPEFSPRLNHNIMNGLSIV--LLLSQGIKVLPELCLFVYFPGK 1527
QY 1584 SUDLAENFLFALSTFTFPDSVAPSGTL-----LATVVALNMG--EL-----D 1629
DB 1528 QKSTLYNN-----GDNDLAVSSSLIKETNQLVLLQKMLNLTANTQELFIQPKNSD 1577
QY 1630 KKKKPLTQAVGSTQAGRTTKSLMFTMENCFYLLISQARKYLDPAVHR----- 1682
DB 1578 DQGEAVELVIGLDSNDKRLSALSKF-----LTHPKYLSRLI 1617
QY 1683 ----DKQMKQELSELSTLSSLSR 1704
DB 1618 PTLLEQQQLDESSRLR-FVKGISR 1642

RESULT 35
Q6C1T5 PRELIMINARY; PRT; 1548 AA.
AC Q6C1T5;
DT 01-OCT-2004 (Tremblrel. 28, Created)
DT 01-OCT-2004 (Tremblrel. 28, Last sequence update)
DT 01-OCT-2004 (Tremblrel. 28, Last annotation update)
DE Chromosome P of strain CLIB99 of Yarrowia lipolytica.
GN ORFNames=YALIOF135199;
OS Yarrowia lipolytica (Candida lipolytica).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
CX Saccharomycetiales; Dipodascaceae; Yarrowia.
NCBI_TextID=4952;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CLIB99;
RG GENOLEYURS;
RA Lafontaine I., de Montigny J., Marc C., Nevegilise C., Talla E.,
RA Goffard N., Frangeul L., Aigle M., Anthouard V., Babour A., Barde V.,
RA Batnay S., Blanchard S., Beckerich J.M., Beyne E., Bleykaesen C.,
RA Boissiere A., Boyer J., Catalicio L., Confantolero F., de Daruvar A.,
RA Despons L., Fabre B., Fairhead C., Ferry-Dumazet H., Groppi A.,
RA Hantrege F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
RA Kerreux A., Kozul R., Lemaire M., Leduc I., Ma L., Muller H.,
RA Nicaud J.M., Nikolaki M., Otae S., Olier-Kalogeropoulos O.,
RA Pellens S., Potier S., Richard G.F., Straub M.L., Suleau A.,
RA Swenne D., Tekela F., Wesolowski-Louvel M., Westhof E., Wirth B.,
RA Zeniou-Meyer M., Zivanovic I., Bojotin-Pukhara M., Thierry A.,
RA Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J.,
RA Wincker P., Souciet J.L.;

```

```

RT "Genome evolution in yeasts.";
RL Nature 430:35-44(2004).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CLIB99;
RA Genoscope;
RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL: CB382132; CAG78184.1;
SQ SEQUENCE 1548 AA; 172980 MW; D1198B6F5FB89E3 CRC64;

Query Match 2.4%; Score 212; DB 2; Length 1548;
Best local similarity 19.4%; Pred. No. 0.00014;
Matches 306; Conservative 240; Mismatches 532; Indels 496; Gaps 78;

213 REOSMLLEIFLYAYFERAPBDLVLTMMKEQGFSGQTRNHLVDETFDPVDRIGYF 272
DB 233 RQVEILKALVNGYFFDFENPSADV-----KQMFESID-NLPFGESBAFC-----Y 280
QY 273 SALIVEGMDISLKA--LDDR--RELHQPADGLICQMDCLMTRPGDIPHH----- 323
DB 281 SILISLVLDLQMFDPVSDLDLDRVDPETDEPIPDNVYFKDPATLTAVANNIIQEHMNEE 340
QY 324 --APVLAWA--LHRTLNPEETS-----VVRKIGTVALQLVN 358
DB 341 LMSPIVLAMSHVLQAVTFROSEETNDRLVPIIDQVSADONGSPYIVCTQLANGSLNRGV 400
QY 359 PGYLRLRLQSLA--SGANDCTSTACMCYGL--LSFVLTSLHLTLANQODIIDTACVYL 415
DB 401 IAVTQSAVAGLAKSAEHTMTWMSLCRAFRVLPVLTAPAEVATVYAGPYGY-----CRYL 456
QY 416 ADPSLPFLFWGTEPFGIILDSV--CGMFPLISPLQLRLALVSGSKTAKKYSPFL 472
DB 457 LENTLK-----IRDMALIASRPFPIRPFVLYMLAM--DGTAFTEVYQWR 501
QY 473 DKM-----SPNELYKHPHDVISHDGLTMRQTKLLYPLGQGTNLRIPQSTGVQWVL 527
DB 502 TLMQQLPGFOXYTESDTNIVTLTGCV-----PVFP-----BFIYPTGSGTLVP 548
QY 528 DDR-AIYRWYYSYSWTLFCIEIMLHVS--TADYIQHQRKPIIDLVHKYIS--TD 583
DB 549 GDNIPPLVMWATREFGMAY-----LQKLSQLSGRFAL-----ITKMRKYISVD 595
QY 584 LSADCLPITSRYMLQLRTTVVSP--VDVIAS-----CYNCLTVLAARPAKY 634
DB 596 AEVADDELRSAS--MMEEDIVYIAKLVDYAYANQHI PVATESKFTALCRIDPDIM 652
QY 635 TDLRHTGFLPVVAHPVSSLSQMSIAGKNNAGYGLNLSNPOQGEYVTIAFLRLITTL 694
DB 653 PLSRAAGIL-----AQNRPGFAAMVIGSTBIIVNGEYSILTALDLADTL 697
QY 695 VKGQIGST-----OSQGLVPCWPFVLEKMLPSYHKMYNSHGVREQIGCLLELIH 745
DB 698 VENNVAASIDSKVSAKESDVIAHCLTRHYI--DVRESCGYMRYVNDQK-----VELMS 749
QY 746 AILMLCHETDLSHSHTSIQFLCISLAYTERAGQTVINIMIGVDTIDMWAAPRSDGA 805
DB 750 KSVAMFOGT-----VHYAPAMONVVPVLA--SAEYIIEGLFTSGE 789
QY 806 EGQGGQGLL-----IKTVGLAFSVTNVIRLK-----PSNVVSPLEQALSQHGAGN 853
DB 790 DLGQRPRLTLCTDSTSEWVNASISLEPLTKIVEARTNLELAPVLEKQFLASPQ----- 843
QY 854 NLIVLAKYIYHKDPLALPRLAIOILKRIATVAMSVYACIGNAAIRDAFLRLQSKI 913
DB 844 -LVELYQGFPHK-----KVIDLMLALVDSSWDCQPSLLAHLGQFHSOL----- 888
QY 914 EDNRKIMLEFLTVAVETQGLTELEFLNLEVKGDSGSKFSIGMNSCLAAVELLDSQ 973
DB 889 -----LKSLSAMEN-----NLE-----TETVIGICAFPAAYVD--GSS 922
QY 974 QQRKYCPPLIARAIAFLALMQRDRDSAMLVLRPKKMENTLSPF----- 1022
DB 923 RQDG-----LSIILVTRKDSILEVLETFALSKQYSSVFPVAVIKAMALAR 968

```

QY 1023 GTLSPSTSEPSILETCALINKICLEYVVKSLDQSL-----KDTLKRFSEKRFAY 1078  
 DB 969 NTWSTVTKTSKORLTK-----LVGKMV--EVVDSVPQLPEKIRSPKSSKRYTVBAACOR 1023  
 QY 1079 MSGVYSKLAHVH-----AETEGSSCTSLLEYQMLVAMPMLLIATTHADIMHMTDS 1130  
 DB 1024 FSA---AAAVQYIABELPRDPTVGA-----LISKW-----GSKMLKLTSS 1062  
 QY 1131 VVRQQLFDVLDGTALLVPAVNCRLGSKCTLLLLLRQ-----WKBELGSV 1181  
 DB 1063 FSSIQLYRSLHG-----NLIRFPAKMMMLLFQFKKDBEYYSFVVDLEDM 1111  
 QY 1182 DEIGCP-LTEILEGVLOADQOLMEKTKAVPSAFITVLOMKMKXSDIFQYSOLV----- 1235  
 DB 1112 NAVFEPSPISGEFSEFAENANLINS-----LVNSSELS--LAWQQLLVAVID 1155  
 QY 1236 LNVCTLOEVALPDQTHSLALG-----SATEDKDSMETDD 1273  
 DB 1156 VKIABSLGDIVALME-----ISLGMLAPVEGVYNNALFVPHILKSPBLKDVYTB- 1209  
 QY 1274 CSRSHRDQDQVGVGLHLAKELCE--VDEGDSMLQVTRRL--PILFTLLTLEVSLEK 1330  
 DB 1210 -----GLKLCWALLESSTVSVLSLSGTSRLKPYLRS-----ISILV 1247  
 QY 1331 KONLHPTA---THLL--LTILARTQCATVAGAGTOSICPLISVYQSLSTGTACT 1384  
 DB 1248 RKALNPEDMAQVHGISDIYLAJ--GSYALANA-----IQYDVSHKG----- 1289  
 QY 1385 PSASRKSIDAPSPGQVYRLSMELQELKTLRYNPLPBEALDFGVYHQBETLLOCLNAVTV 1444  
 DB 1290 --HSEITVDVDFLIEDLRIMVALLRNL--SFENTCVPSVLAQAL---KIDDTGSIRSL 1342  
 QY 1445 QSL-ACLEBAHTVGFILQLSNFMKEMHFLPOLMRDIGNVLYLCOACTSLHRSKMLQ 1503  
 DB 1343 LSLVYSQGEIDPVFG-----ELTLLFILE--LLQNDQJSE 1375  
 QY 1504 HYLONKNGDGLPSAVAQVQPPSAASAPSSSKOPADTEASBOQALH--TVQY---G 1557  
 DB 1376 QVVLN---GLVGAALTE-----APISRE-----IQKSLHQGLHYIWMIRG 1412  
 QY 1558 LKLIKSLTAL-RHFTPVQCLLDQSLDLAEYNLFALSTPTTPPESEVAPSEGTILA 1616  
 DB 1413 ILPIMLYLVRLGGRISPEITML-----LSF---FOUQIETALRNLME 1452  
 QY 1617 TVNVALNMLGELDK 1630  
 DB 1453 PVTISLSFVDELDQ 1466

RESULT 36  
 O6CNPL PRELIMINARY; PRT; 1606 AA.  
 AC 06CNPL: 01-OCT-2004 (TREMBlrel. 28, Created)  
 DT 01-OCT-2004 (TREMBlrel. 28, Last sequence update)  
 DT 01-OCT-2004 (TREMBlrel. 28, Last annotation update)  
 DE Strain NRRL Y-1140 chromosome B of strain NRRL Y-1140 of Kluyveromyces  
 lactis.  
 GN ORFNames=KLLAOE11033g;  
 OS Kluyveromyces lactis (Yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.  
 OX NCBI\_TaxID=28985;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NRRL Y-1140;  
 RG GENOLEVURS;  
 RA Lafont B., Sherman D., Fischer G., Durrens P., Casaregola S.,  
 RA Lafontaine I., de Montigny J., Marck C., Neugebilde C., Talla B.,  
 RA Goffard N., Frangoul L., Aigle M., Anthouard V., Babour A., Barbe V.,  
 RA Barnay S., Blanchin S., Beckerich J.M., Beyne E., Bleybaesen C.,  
 RA Boistame A., Boyer U., Catolico L., Contalioletti F., de Daruvar A.,

RA Despons L., Fabre E., Fairhead C., Ferry-Dimazet H., Groppi A.,  
 RA Hantreay F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,  
 RA Kerrest A., Kozul R., Lemaitre M., Lesur I., Ma L., Muller H.,  
 RA Nicoud J.M., Nikolaki M., Ozias S., Ozier-Kalogeropoulos O.,  
 RA Pellenz S., Potier S., Richard G.F., Straud M.U., Suleau A.,  
 RA Swenne D., Tekela F., Weselowski-Louail M., Westhof E., Wirth B.,  
 RA Zentou-Meyer M., Zivanovic I., Bolotin-Fukuhara M., Thierry A.,  
 RA Bouchier C., Caudron B., Scarpelli C., Gallardin C., Weissenbach J.,  
 RA Winkler P., Souciet J.L.,  
 RT "Genome evolution in yeasts",  
 RL Nature 430:35-44(2004).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NRRL Y-1140;  
 RA Genoscope,  
 RL Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.  
 DR EMBL, CR382125; CAG99535.1;  
 SQ SEQUENCE 1606 AA; 182208 MW; 0C5837B000AF2363 CRC64;

Query Match 2.4%; Score 212; DB 2; Length 1606;  
 Best Local Similarity 17.4%; Pred. No. 0.00015;  
 Matches 339; Conservative 295; Mismatches 606; Indels 712; Gaps 90;

QY 76 KELGRIKSGLDDEQSVOLQCTLOEDYGRDSDVKTVLQDERQOALIKIADYYE 135  
 DB 38 EELVHLSRFTSVENTDSVG-----KOLQNGY--SYVATVEDELAMBLIALNVHQS 90  
 QY 136 -----ERTCLRCVLLHLYTPDERRPRVREYVADCYDKLEKELVSKR 178  
 DB 91 CLRVISQTRTFPPDCKPMKDTIVKQILH-----ER-----NCV--LDITVLNLH 133  
 QY 179 QQFEL-----YKTEAPTWETHGNLMTREV-----SRMFOQLRQSMLEIIFL- 224  
 DB 134 DSEFVLKSMQOLVYKNRS--TWISAILLVKNCVALISABQNDYSKLSYDILYIN 192  
 QY 225 ---YAAVEM-APSDLVLTLMFKKEQSGRQTNHLYDETDMPVDRIIGVFSALILVG 280  
 DB 193 LNLNLPFVSLNQVLELITKYE-----LLEFQ-----YEA--ILKDN 230  
 QY 281 MDISLHKCALDRRELHQAQDGLICQDMCLMT-----FGD----- 319  
 DB 221 KDVP-----EIVMANSLMI--MDIIMTYNTSBDITDVKSPFNDAATFR 277  
 QY 320 -----LPHAPVLA--WALL-----KHTLPEETSSVVRKIGTA 353  
 DB 278 LLDLIINDPPALIVYVYSPILYKSYLLEEDPENNLFPVQOVFNSQIPISKIKFALTA 337  
 QY 354 IQLNVFOYLTRLQSLASGNDCTSTACMCYVGLSFVLTSLEHTLGNQDIIIDTAC- 412  
 DB 338 ENENVFERSKVSQLF-----KQAFPAALILSFITPSLYFVPISIKISMITVLSNAPT 393  
 QY 413 ---EVLADPSLPELFWGTEPTSGILIDSVCGMPLHLSPLLRLALVSGSKTAKV 468  
 DB 394 HYVEKFLNKEPEKFF-----LIKTKAPLKESLIPMINITS- 432  
 QY 469 YSFLDKMSFVLELYHKP-----HVISHEDTGLMRRQTPKLY 507  
 DB 433 ---NAEFANFEMWELPTVYQEVVLSLELDLADASTDILSIKSETPA- 480  
 QY 508 PL--GGQTNLRIFQSTGVQVM-----LDDRAYLVREYSGYSWTLF-----TCSEMLH 555  
 DB 481 FLBPSGNVLPFIPKQTKAKILPSVIND----VMLELYKXNGSLGRLVQNCND-----R 532  
 QY 556 VVSTADV-IQHC---QRYKPIIDLVHXYISTDLST---ADCLLPTSRITMQLRLTYI 608  
 DB 533 YLSTSGAVQDSVDQEMILSMDLISSVYSGISVEKADILLQNST--YLTEDYIFDI 590  
 QY 609 SP-----PVDVYACVNCGLTTLAARNPAKVTDLRHGFLPFVAPVSSLOMISA 659  
 DB 591 FVYFQSLHRLADYDIDICLNLILSSIDDPSQFVSHL-----ARSLDLR 636  
 QY 660 BGMNAGYGNLNMSEOPQGBYGTIAFLRLITTV-----KQQLSGTQSG 705

Db 637 YG-KSGLISSVLGTMELPENGNYITIKIKLAQOVLSESLTETEPENIKKELISKPFQ 695  
 Qy 706 GLVPCWFWLKEMLPSYHKMYNSHGVREJGCLILBILHAILMCHETDLSHTSLO 765  
 Db 696 HFI-----HVESIQYNNYRSIAEKELGTMLTTLFTVNNIGID--PSNPETK 745  
 Qy 766 FLCISLAYEAGQVINIMIGIVDTIDMWAAPRSDAGGQOG-----QLIK 816  
 Db 746 ITKVLSSAANDITNRFLSSDSPOVRTYSILHVLSSNSLSALSGNEVPFNKNOCLIK 805  
 Qy 817 TVKLASVTNNVIRK-----PSNVVSPLEQALSQGHAGNNLIIVLAKYIKHHPAL 871  
 Db 806 SYELA-----NLLSVRAMLOLPSTLESQI-----YGN-----L 835  
 Qy 872 PRLAQLKRLATAPNSVACGNDAAIRDAFLTRLOSGLIEMRIKIMILEFTYAVE 931  
 Db 836 PKL-----IDITYA-LHND-----LRNPI-----KIMTHLVRAWAB 866  
 Qy 932 TOPGLIELPLNLEVKDSDGSKFSLGMSCLHAHLEIDSQODRYWCPFLHRAIAF 991  
 Db 867 NPSILSLYL-----GQKHSMKMFNSILYDLASTIDYQ-----PLKSLYRF 907  
 Qy 992 LHALMODRDSAMLVRT-----KPKFWENLTSPLFGLSPSETSPSILETCALI 1043  
 Db 908 PSSVMEGKODGFSILVTGETISVSGSPKSE-----PSNGASSTIL--SL 953  
 Qy 1044 MK-IICELIYVVVKGSLDQSLKOTLKPFSEKRPAYMSG-----YKSLAVHAYET 1093  
 Db 954 KKNVLKIDCY-----PESVASRLLE-AISYANSAARANNATADKEFIDFVKLEKF 1005  
 Qy 1094 EGSSCTSLLEYQW-LVSAMRMLLIATTHADIMHLPDVSVRQLFDGTALILVPA 1152  
 Db 1006 EDLAPDSVEATIKLSGHRVLOABIAEIFA--LHLPSSPSDSTIMMANGPO---LVDI 1060  
 Qy 1153 SVNCLGSLGKCTLLILKQMKRELSVD-----EIL--- 1185  
 Db 1061 VKRIFIDGYDKTILHLEEFKMIWPHLDIEKPAISPLRIGKAYQANIYDIELDQF 1120  
 Qy 1186 -----GPILEIGVLOADQOLMEKTKAVFSAFITVL---QMKMKYSDIP 1229  
 Db 1121 CCEKRGKNGSTGFEQVKSINLOYSFQISAASWGALLTAPTKTPKLOASVYD 1180  
 Qy 1230 QYSQV-LNVCEFLQEBVIALFDQTRHSLALGSATEDKOSMETDSCRSRHRDQRGVCY 1288  
 Db 1181 IVTQLALNVHEGVAPIFS-----EIVL 1204  
 Qy 1289 LGHLAKECEVEDDQSWLQVTRRLP--ILPTLITLLEYSLEMKO-----NL 1334  
 Db 1205 LRLELVFYLY-----SFLQTSMKIPEDLETLISQLALILKSKEVDYLGNIAGNKN 1258  
 Qy 1335 HFTL-ATLHLTLTARQOGATAVAVAGITGICPLSVQSLSTNGTAQPPSARSL 1392  
 Db 1259 QVRPIRLSILMLITLTKPY-----PMFSEERS 1285  
 Qy 1393 DAPSWGVRYSLSMT-MEQLKTLRYNPLPEALDFVGHQERTLOCLANVRVQSLACT- 1450  
 Db 1286 D-SLEVEPELAFSSKGVNIIISILSPFT-----IASNKGKVIVPNLADKIDQLLLL 1337  
 Qy 1451 -----EEDHVTGVPFLQSLNFKEMW-----FHLPOLMD-----IOVNIQYLQ 1490  
 Db 1338 SLETKIKELNPSSTSFNAVILASSINBEGTLKSLNLTLYSSYLFPSHDEPLFAELTSLFIB 1397  
 Qy 1491 ACTSLHSHKMLOHVLOONKNGOLPSAVAVQVRQPPSASAPSSKOPRAOTREASEQA 1550  
 Db 1398 LVT-----IEEVAQKIINDGLFSVPLE-----SPISVSIQSGAPPELSPR-- 1438  
 Qy 1551 LHTVOYG-----LKLISK-----TLALRHPTD 1575  
 Db 1439 LANINSGIISIILOLSKFGSGLSEVCLFVSYFESKOJDLAVGMSSSLITQAYIAB 1498  
 Qy 1576 VQO-IILDDSLDLAEYNPLFALSFTTPTDSSEVAFSGTLLATVNAALMLGELDKKCP 1634  
 Db 1499 TSOIILQKMLDKLEQRYLSTVN-----GSKLKYID-----EERI 1533

Qy 1635 LTOAVGLSTOAG-----TRTLKSLMFTMENCFFYLLISQAMRYLRDAVHR----- 1682  
 Db 1534 IELPGLDYSSEKNDLANVTKHL-----THPKYNSRIIAT 1570  
 Qy 1683 --DKORWKOB-----LSSELSTLSL 1702  
 Db 1571 TPDKQKMEDEYVRNQLHLSITQISLQSSSL 1602  
 RESULT 37  
 ID 09LUS3 PRELIMINARY; PRT; 1837 AA.  
 AC 09LUS3;  
 DT 01-OCT-2000 (Tremblrel). 15, Created  
 DT 01-OCT-2000 (Tremblrel). 15, last sequence update  
 DT 01-DEC-2001 (Tremblrel). 19, last annotation update  
 DE Similarity to unknown protein.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
 OC euroside II; Brassicales; Brassicaceae; Arabidopsie.  
 NC NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20181125; PubMed=10718197;  
 RA Sato S., Nakamura Y., Kaneko T., Katoh T., Asanizu B., Kotani H.,  
 RA Tabata S.,  
 RT "Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence  
 RT features of the regions of 3,076,755 bp covered by sixty P1 and TAC  
 RL clones.";  
 DR DNA Res. 7:31-63 (2000).  
 DR EMBL=AB023044; BAA97383.1; -  
 SQ SEQUENCE 1837 AA; 206399 MW; 45BB6A7B0E70795 CRC64;  
 Query Match 2.3%; Score 205; DB 2; Length 1837;  
 Best local similarity 18.7%; Pred. No. 0.00056;  
 Matches 377; Conservative 273; Mismatches 708; Indels 662; Gaps 93;  
 Qy 48 LKGLSYKPPSPSSAKRYKA-----NKDVASPKELGLRISKFLGDEQS 93  
 Db 40 LQNLSPFPKPSDQAQVQSKERLPLSLDQDIA-----ISLKLSDLEHNLNRIDS 94  
 Qy 94 VOLLOCVIOE-----DYRGTRSVTVLODERGQSLIKIDVYBS 135  
 Db 95 VRLVSSNOEGMGRDPLEIORLATGLMYTRRDLTSLY----- 135  
 Qy 136 ERTCLRCVLMILTYFODERHPYRVEVADCVDKLEKELVSKYQOFEBLYK---TEAPTW 192  
 Db 136 ---TLRAVY-----LDGGLBBDLI--ADIGLLEELIENAGLRQRLITLIKELANREDPTG 185  
 Qy 193 -----ETHGNLTEROVSRWFVQCLREQSMLEIIFLYAYAYEMADSDLLVLT 241  
 Db 186 LGGPLCEHYLLIDSRGALVERBAV---VQ---RERLILGHCVLSILVDRPGSKVDVDIY 239  
 Qy 242 MEKEGFGSRQNRRLNVETMDPFYDRIGY---PSALLVSEMDLSLHKCLDDBREIH 298  
 Db 240 ILKD-----NAAQLTEGNDTISQITPFLSLFLITTPSDALS-----RLSDKSSN- 285  
 Qy 299 QPAQDGLICQDMCMLTFGDIP---HHAPVILAMALLRLTINBERTSVVRKIGTAT 354  
 Db 286 -ISQDASFRTPQDILVMASSGSDPTADGRTGGIRLAMAN--HLM----- 325  
 Qy 355 QLVNFQYVTRLLQSLASGANDCTSTA-----CMCVYGLLS-----FVLTSLELHTIGN 403  
 Db 326 -----LIHQISGMOTITSTASTDWMGHICGLSEISPKVNFQFLDNV--LRTAY 374  
 Qy 404 Q-QDIIIPACVYL---ADP-----SL-----PELPKGTBPTSGIGI 435  
 Db 375 QVKSCKDNAMSVLNSYRTSDPLDGSMTGERSDRPLPFLISLMEFKPELISGND----- 427  
 Qy 436 ILDSVCGMPPHL-----LAPLLQLALAVSGKSTAKKVVGFLLKMF-- 477

428 -----VLMTFVNPAEDHTNFKTLVAFLEMLCTLASTOGASKVYELNGTSPRSIGMPTL 483  
 QY 478 -----YNELYKHKPHDVIHSHEDGILMRQPKLIYPLGGQTNLAIPQGTGVQVNLDR 531  
 Db 484 FDCIRIDEKFKQS-----LQTAGAMMP-----BFLGDAK 514  
 QY 532 YLVMEVYSASWTLFTCEIMLHVSTADVIHOCORVPIIDLVHAKYSTDLADIADGL 591  
 Db 515 ALVAVY-----LNVLOKVENGNTPTER-KAMFPDIEPPFKLGIE----- 552  
 QY 592 PITSRIYMLQRLTTVISPPV-----DVASCVNCLTVLAARNPAKWTDLRHTGFLPF 645  
 Db 553 -----NIPYLKALRKTIAPFVNVPEMR-----DSIWAFLQYDLPV 592  
 QY 646 VAHVPYSSISQMIASGNMAGVGNLNMSEOPQSGYTIAPLITLVKQGLSTQSQ 705  
 Db 593 VGSQVQKSDQSOVDMQFE-----LNEVEAREQYPSITISPLINLNLAGE-KDVNDR 646  
 QY 706 G-----LVPCVWFVKEMLPYHMKRYNKGHVRBOJGCLILELHAILNL--CHETDLH- 757  
 Db 647 GRRAVSDPC-----EKM-----QLVAVACLOHFMHLSMTYDIOEDLDGF 685  
 QY 758 -----SSHPTSPLOF-LCICISLA--YTEAGQTVINIMGI---GVDTIDMVMMAQPRSD 803  
 Db 686 TEHPFLVSLFTSSLOQOLPIIILKDFMSSKALYRNLMLQVGVNSIISERLSKT--- 742  
 QY 804 GABGQGGQGLIKYKAFSVATNNYIRLK-PPSNVVSPLBQALSGHAGNNLIVLAKY 862  
 Db 743 -----YGKILEKAVQSLLEILLVFEKOLLVSDWRBLYOPLDIILQDHOIALLLEY 796  
 QY 863 IYHGDPALEPLATOL-----KRLATVAPMSVYVACLDNDA-ARDAFLTRLOSKIEDM 916  
 Db 797 VRVDSLPQIQSSIKIMNILRCRSLVGLVPMILKI---DAANSILBYAACLSERLEBG 852  
 QY 917 RI-----KXMLEPLTVAV-ETOPGLIBFLNLEVYDSDGSKERSLGMWCSLAHVL 967  
 Db 853 EVVENSODDLGLVLMQILVNDINRPAISITHLILFDDAPBEGVLPKPHYSGLKXIL 912  
 QY 968 ELISQOQDRYWCPR--LHRAIAFLHALWQDRDS--AMLVLRTPK--FWENLTP 1020  
 Db 913 EMELKLPY-----PDINFLPEFGFQOLCEMLDPLTGPTWDLSSKKYOFLOHLDLT- 966  
 QY 1021 LFGTLPSP-----SESEPSILETCLIMKITCLBY-----Y 1053  
 Db 967 -IGVATLPKRSQSLARISLSHQRAWMLKLAIALHTSGSSSAHLKACQSLSLHFERE 1025  
 QY 1054 VVKGSLDQSLKDTLKKFSIEKFAVWSGVKSLAVH-----AETEGSSCTSLIE 1103  
 Db 1026 VTEBANERFSSSTYQDGLD--YAGTSSISKSKALALEILOFRPDSMQLPOIVSSLK 1083  
 QY 1104 YQMLVSAWRMLLIATTHADIMHLLTDSVVRQLF-----LDVLDGTAKL-- 1148  
 Db 1084 YDSLYE-----DILGNRDTVSQSGIYYERSGRDLIDLSFSPNLMQKLSHG 1130  
 QY 1149 --LYPASVNCIARLSMKCTLLIILIRK-------RELQ 1179  
 Db 1131 PPLVDSFENVALSEVERETIOQLKMGKYNRNLEQAOLHMLAGSQIVEVASCRIS 1190  
 QY 1180 SVDELIGLUTRILBGLVLOADQOLMEKTKAVPSAFI-----TVLOKMEKMS- 1226  
 Db 1191 SLDN--RSEILIRILIDNASASASAPCSLMALVULOVALTCIAKARDRFRSQALS 1246  
 QY 1227 -----DIPQYQVLNVV-------ETLOEVAL--FDQTRHSLA 1258  
 Db 1247 SDVTYCLDVMVVKHLSGTAGCHSVFLKVMALIRHSSSSLRARQVALLSYQYQCHMIA 1306  
 QY 1259 LGSAT-----EDKDSMETD--DCSSSRH--RDQDQVC- 1287  
 Db 1307 LDVPTSVVQFLNLNODGEDDLDIOIKIDKQADLABANFPIIKKAGGILDLVYKASQGS 1366  
 QY 1288 -----VLGHILAKELCEVEDDQSMLOVTRRLPIFLTLTLEVS-----LRMKONLHF 1336  
 Db 1367 EFGKTIISLYVABALVCIDHERRYFISQLOGRPIRSCLSISINISYODGTHLLBSQORACT 1426

QY 1337 TEATLHLLTLART--OOGATAVAGAGITOSICLPLSVOLSTNGTAQTPSARSKSLDA 1394  
 Db 1427 LEAEIALMLRISHKYGSGGQVLPMSGALBHIA-----SCRAISFKGNKR-----RVDMKL 1477  
 QY 1395 PSMFG-----VYRLSMIMEQLKTL-----RYNLPBALDPVGHQERTLOC 1437  
 Db 1478 QSDVGVNVQKQRIITIVLRVLPAL-TSLVETSEFPERRNKIYADVAFIKGHQGLPQL 1536  
 QY 1438 LNAVTVOSLACBBAHTVGFILQLSNPKEMHFLPQLMRDIOVNLGYICQACTSLIH 1497  
 Db 1537 LREDFTQADMLMEQIILAVGILSKWMPFEENDGQFVQGIQFDMWSKL-FIASPIKSIYS 1595  
 QY 1498 SR-----KMLQ-----HYLOKNG-----DGLPSVAORVORPPSAASAPSSK 1537  
 Db 1596 QKSSBLKLSQRFSLTYLTVTKNSLRLOVSDSLDST--KIROPTLLILSLSHV 1653  
 QY 1538 QPAADTEASEQC-ALHTVQ-----YGLKILSKTLAALRHFP- 1574  
 Db 1654 TDSLEBAEKSLHKKRIINELSRQVDNALIKICD--SQEYVTPSDIHKKRYIAM 1709  
 QY 1575 -DVQOI-----LIDQSLDLAEVNP-LFALSFTPTPDSVAPSEGTILATVVALNMG 1626  
 Db 1710 VEMCOIVGNBDQTLTLQLLAHVLIINILHODRSVSSNERGSGSKSHIQOEVTDLCG 1769  
 QY 1627 ELDKKKEPLTOAVGLSTQABGRTLKSILMFTMENCPLL 1666  
 Db 1770 KL-----SPTIDRLALNEV-----SLIDREPSIIVL 1798

RESULT 38  
 OTTIS3  
 ID 071153 PRELIMINARY; PRT; 1278 AA.  
 AC 071153;  
 DT 01-OCT-2003 (TREMBLrel. 25, Created)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last sequence update)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
 DE SI:ZC146F4.2.2 (Novel protein similar to human BAR28) (Fragment).  
 GN Name=SI:ZC146F4.2? (zebrafish) (Danio rerio).  
 OS Brachydanio rerio (zebrafish) (Danio rerio).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Butelosteomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Osteichthyes; Cypriniformes;  
 OC Cyprinidae; Danio.  
 OX NCBI\_Taxid=7955;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Garner P.;  
 RL Submitted (JUL-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AL732629; CAB17602.1; --  
 FT NON TER 1  
 SQ SEQUENCE 1278 AA; 143523 MW; 86C9FC81B77EB1A9 CRC64;

Query Match 2.2%; Score 197.5; DB 2; Length 1278;  
 Best Local Similarity 18.4%; Pred. No. 0.001;  
 Matches 241; Conservative 180; Mismatches 401; Indels 487; Gaps 59;

QY 677 PQGEVGTIAFLRLITLVKQGLSTQSGVPCVMFVLEKMLPSYHK----- 724  
 Db 6 PEKMETTCCYLRILCRFLFVAVIGASQGRAPCFRSLMOPLOVHLNEPVLFFKFLSL 65  
 QY 725 WRVNSHGVRBOJGCLILELHIA-----LNLCHETDLSSHSTPSLOPIC 768  
 Db 66 WGNYSN-LGDDLDCKVSAIILQTOALYGKAPLSQPVKTLMLASDS--SPVVSBL-LVC 121  
 QY 769 ICSLAYTEAGQTVINIM-----GIGVDTIDMVA--A 798  
 Db 122 VCS-GVCEVRRALAVAVLOCLSGIVSSPYHPLVEKLKSSSEKILIDSSVLTQALSFYBEA 180  
 QY 799 QPRSDAGGQGGQGLIKTVKLAFSVTNNVIRLKPSPSNVVSPLBQALSOGHAGNNLIV 858  
 Db 181 VSRDKKNKGLASVQLOCLQSPFC-----PSYTSKTLRALOD--VHGEPVLIV 228  
 QY 859 LAKTYHNDPALRRLAIQLKRLATVAPMSVYVACLDNDAALRDAFLTRLOSKIEDMI 918



```

Db 229 LL-----PAYERLLEQCAPDSCFLP-----DEALL-----LOLLLSKFSMSA 267
Qy 919 KWMILEFLTAVETOPGLIELFLNLEVKDSDGSKFSLGWSCLHVALELIDSOQDRY 978
Db 268 FLVVKD-----PRCLEVFIR----- 282
Qy 979 WCPPLHBAIAF-----LHALMODRDSAMLVLRKP-----KFMENLTSPF- 1022
Db 283 ----ALHTSARPYPTIPSFQITALEQ-----ITKFPFALIGDEKIQOKILSIFD 328
Qy 1023 ----GTLSPSEETSEPSILETCALIMKICILEIYV-----YKGSUDOSLKDITLKFSIERK 1075
Db 329 LTVGNKSPACAGINSVFKTIAVDCELVANELLIPADKORVATATVQOTRSKRRK----- 382
Qy 1076 FAYMSGYVSLAHVAETBSSCTSLLEYOMLVASAMRMLLIATTHADIMELTDSVVRQ 1135
Db 383 ----TODTSGA-----VPBESVSWPRVTLIL-----ELLQHKKGLKRAQ 418
Qy 1136 LFLDVLDTKALLLVPSAVNCLRGSKCTLLILILKORMBELGSVDEILGPLELLEGV 1195
Db 419 YLVPLFNLISRCLEPAAAGENIETKOLILICLAVCOK-----LSPGGPIS---KDV 471
Qy 1196 LQADQOLMEKTKAKVPSAFITVLQMKEMKVSIDIPO----- 1230
Db 472 LEBDKFNME-----LVQCVRSMPQTHHALLLGLAGIFPEKVLANIM 518
Qy 1231 ----YSQVLAVNCETLOEBEVALFDQTRHSIALGSATBEDKDSMETDCC 1274
Db 519 PIFTEFGANIMRLDDTYSFOVIN--KTIVQAVIPALI--KANE---GSSSOSBGMET--- 568
Qy 1275 SRSRRDQDGVCVLGLHLAKELCEVDEBODSMLQVTRRLPILPTLTLEYS----- 1327
Db 569 ----VVAQIIVFDALPVPFHRRLPILSOLMSTIGSPERFLVIM 610
Qy 1328 -LRMKNHPTENTALLTLARTOQATAVAGAGITOS-----IC----- 1367
Db 611 LLLFQO--HVTQ-----TSAGATGAKBAVVERDQDFWILVCCFEYKBOJL 657
Qy 1368 LPLSLYSTGNGTAOTP-----SASRK-----SLDAPSPGVTR---LSMGL 1407
Db 658 IKILDYIMTLPODREBAPEKRRGRSAVKDETVSDLIFSVEHSGKLRHFKFISISF 717
Qy 1408 MEOLKTLKLYNPLREALDFVGNHOBERTLOCL-----NAVPTQOSA--CLEE-AD- 1454
Db 718 MAQOLAS--DSFVGKADCEDI--TESTLOALQODLVEVLRKYOAVARCVEDNADKFTAK 774
Qy 1455 ----HTVGFILQDSNFMKEWHFHLPOLMRDIOVNLGYLCOACTSLHSRKM 1501
Db 775 FWRALLSKSYDTLDKVNALLPMDTFI-----TMRGLMGN-----QLASVRK 817
Qy 1502 LOHYONKRGDGLBSAVAVORVORPSAASABSSSKQPAADTBASQOLMHTVOGLKI 1561
Db 818 AMELLNKK-----LQOR-----TKMKKEQIT--ALLET 843
Qy 1562 LSKTLAAL-----RHFTPDVCQILIDOSLPLAEYNFLFALSFTTPPDSSEVARSFGLLAT 1617
Db 844 IGTLLSIYGRSHROQTAGE-----EBELAINQTAIYSLKLCRNFGSDHKEEF---VFV 895
Qy 1618 VNVALNMLGEIDKKEPLTOAVGLSTOAGRTLKSLM-----FTM 1659
Db 896 LNKAVELVADKDEBKVNGSA--LLCVAVSTLTKALAIPOHRLMPAVLDTLKRRKOLL 953
Qy 1660 ENCFILL--ISQAM-----YLRDPAVHRDKQKMQOELS--ELSTLLSSISJR 1704
Db 954 NNEIYLLSAVVALQASSETLPHRISPYLLDTLOVTRTLARRLTSCQSLSVRLASISS 1013
Qy 1705 YFRGAPSPATGVLPSPQGSLSKSPESQEPILQIYQAFVHMR 1753
Db 1014 TL---ATKLPRAVLPIITTKYCSWVDQOQNLSPMLNKHSHMK 1059

```

RESULT 39

```

07T152
ID 07T152 PRELIMINARY, PRT, 1336 AA.
AC 07T152;
DT 01-OCT-2003 (Tremblrel, 25, Created)
DT 01-OCT-2003 (Tremblrel, 25, Last sequence update)
DT 01-OCT-2003 (Tremblrel, 25, Last annotation update)
DE ST:ZC146F4.2.1 (Novel protein similar to human BAP28) (Fragment).
GN Name=ST:ZC146F4.2;
OS Brachydanio rerio (Zebrafish) (Danio rerio).
OC Actinopterygii; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Cyprinidae; Danio.
OC NCBI_Taxid:7955;
RN [1]
RP SEQUENCE FROM N.A.
RA Garner P.;
RL Submitted (JUL-2003) to the EMBL/Genbank/DBJ databases.
DR EMBL; AL732629; CAEL7603.1; -.
FT NON TER 1
SQ SEQUENCE 1336 AA; 150326 MW; AA949557P21ACBCE CRC64;

Query Match 2.2%; Score 197.5; DB 2; Length 1336;
Best Local Similarity 18.4%; Pred. No. 0.0011;
Matches 241; Conservative 180; Mismatches 401; Indels 487; Gaps 59;

Qy 677 PGEYGVITAFRLITTLVKGQSGTOSGVLVPCWFLKEMLPYHK----- 724
Db 6 PERMETTCCYRLRLCRLEFDVVISGASQGPLACFRSLMQPLQVHNEPVLFFLSLL 65
Qy 725 WRNNGHVRBOIGCLILEIHA-----ILNCHETDLSHTPSLQPLC 768
Db 66 WQYNIN--LGDQDCVSAALQTOALYVGAFLSPQVKTILMADS--SPVPSL-LVC 121
Qy 769 ICSLAAYTAGQTVINIM-----GIGVDTIMVNA-----A 798
Db 122 VCS--GVCEVRAALAVLQCLSGLVSPHYPLVEKLKSSBEILADSVLTQLSKYBEA 180
Qy 799 QPRSGAAGGQGGOLITVTKLASVYNNVIRLKPSPVNVSPLEQALSOHAGNNLAV 858
Db 181 VSRKDKNNKLASVEBLOCLQSPFC-----PSYTSKTLRALD--VHGSPVLSV 228
Qy 859 LAKYIYHNDPALPLAIOQLKRLATVAPMSVYACGNDAAIRDAFTRLOSXIEMRI 918
Db 229 LL-----PAYERLLEQCAPDSCFLP-----DEALL-----LOLLLSKFSMSA 267
Qy 919 KWMILEFLTAVETOPGLIELFLNLEVKDSDGSKFSLGWSCLHVALELIDSOQDRY 978
Db 268 FLVVKD-----PRCLEVFIR----- 282
Qy 979 WCPPLHBAIAF-----LHALMODRDSAMLVLRKP-----KFMENLTSPF- 1022
Db 283 ----ALHTSARPYPTIPSFQITALEQ-----ITKFPFALIGDEKIQOKILSIFD 328
Qy 1023 ----GTLSPSEETSEPSILETCALIMKICILEIYV-----YKGSUDOSLKDITLKFSIERK 1075
Db 329 LTVGNKSPACAGINSVFKTIAVDCELVANELLIPADKORVATATVQOTRSKRRK----- 382
Qy 1076 FAYMSGYVSLAHVAETBSSCTSLLEYOMLVASAMRMLLIATTHADIMELTDSVVRQ 1135
Db 383 ----TODTSGA-----VPBESVSWPRVTLIL-----ELLQHKKGLKRAQ 418
Qy 1136 LFLDVLDTKALLLVPSAVNCLRGSKCTLLILILKORMBELGSVDEILGPLELLEGV 1195
Db 419 YLVPLFNLISRCLEPAAAGENIETKOLILICLAVCOK-----LSPGGPIS---KDV 471
Qy 1196 LQADQOLMEKTKAKVPSAFITVLQMKEMKVSIDIPO----- 1230
Db 472 LEBDKFNME-----LVQCVRSMPQTHHALLLGLAGIFPEKVLANIM 518
Qy 1231 ----YSQVLAVNCETLOEBEVALFDQTRHSIALGSATBEDKDSMETDCC 1274
Db 519 PIFTEFGANIMRLDDTYSFOVIN--KTIVQAVIPALI--KANE---GSSSOSBGMET--- 568

```



QY 1275 SRSHNRDQRCVGLIHLAKELCEVDEDDSGMTQVTRRLPILTLTLEVS----- 1337  
 DB 569 -----VAAQIHVFYDALPHVPERHRLPILSOLMSTLSPSFLWTLM 610  
 QY 1328 -LRMKQNHFEATLHLLTLARTQGNATVAGITQS-----IC----- 1367  
 DB 611 LLLFKQ--HTQ-----TSAGTGAKEAVVERDDFWILVCCERFVEQLTSL 657  
 QY 1368 LPLISVYQSTNGTQTP-----SASRK-----SLDAPSPGVTYR---LSMGL 1407  
 DB 658 IKILQYMTLPDREBAPEKKPRGRSAVKDETVSDILFSEVTHSGDLMHFRFISISF 717  
 QY 1408 MEOLIKLRNLEPDLDFVGHQERTLOCC-----NAVTVOSLA-CLEB-AD----- 1454  
 DB 718 MAQULAS--DGFQGVADCEDI--TESTILOAQODLLVEVLTYQVACVEDNADKPTAK 774  
 QY 1455 -----HTVGFILQLSNFKEMHFLPOLMEDIOVNLGYLCOACTSLHSRKM 1501  
 DB 775 FMRALLSKSYDTLDKNAALPMDTFI-----TWKGLMGN-----QLASVRK 817  
 QY 1502 LOHYLQKNGDGLPSAVAQVQRPSPASASAPSSSKOPADTEASEQALHTVOYLKI 1561  
 DB 818 AMELNNK-----LQQR-----TKWLKEQIT-----ALLEL 843  
 QY 1562 LSKTIAL-----RHFTPDVQIILDDSIDLAENYFLALSTFTTPDSEVAPSGTILAT 1617  
 DB 844 IGTLLSTVGRSHRQVTAOE-----EELAIRQRTALSLKLLCNFGSDHKEBF---VPV 895  
 QY 1618 UNVALNLGELDKKEPELTQAVGLSTQAEGRTLKSLIM-----FTM 1659  
 DB 896 LNKRAVELVADKDEKKNWGS--LLCVAFVSTIKALALPOLHMLPAVLDLTKERDOL 953  
 QY 1660 ENCFYLL--ISQNR-----YIRDAVPRDKQKQKQELSS--ELSTLSSLSR 1704  
 DB 954 NNEIYLLSAVTAQRASFTLPHFISPYLDITLLQVTRLLTARLITSCPOLSVLASISS 1013  
 QY 1705 YFRGASSPATVLPSPQSGSTLSKSPSPOEPLQVQAFRHMQR 1753  
 DB 1014 TL---ATKLPFRVLIPTTKCYCSMVDAQNRSLPMLNKLKHSMDK 1059

RESULT 40  
 QYVMB8 PRELIMINARY; PRT, 2067 AA.  
 AC QYVMB8;  
 DT 01-MAY-2000 (TREMblrel. 13, Created)  
 DT 01-OCT-2002 (TREMblrel. 22, Last sequence update)  
 DT 01-OCT-2002 (TREMblrel. 22, Last annotation update)  
 DE CG11943-PA.  
 GN ORFNames=CG11943;  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OC NCBI\_taxid=7227;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Aame M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherzer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.H., Blazer R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gaber G.L.,  
 RA Bailey J.F., Agbayan A., An H.J., Andrews-Pfankuch C., Baldwin D.,  
 RA Ballew R.M., Baer A., Baxendale J., Bayraktaroglu L., Beasley B.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Botkova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,  
 RA Butte K.C., Butman D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Palacios B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.B., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Fierla S., Fleischmann W.,

RA Foeller C., Gabrielian A.B., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodde A., Gong F., Gorrell J.H., Gu Z., Guan P., Harrie M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Mei M.H., Ibegwan C.,  
 RA Jatali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Kechum K.A.,  
 RA Klumel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Matcel B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Mekulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Pacle J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen T.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier B., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodruff, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,  
 RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zhang X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
 RA "The genome sequence of Drosophila melanogaster."  
 RL Science 287:2185-2195(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426065; PubMed=12537568;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J.W., Halpern A.,  
 RA Patel S., Adams M., Champe M., Dugan S.P., Frise B., Hodgson A.,  
 RA George R.A., Hoskins R.A., Laverly T., Muzny D.M., Nelson C.R.,  
 RA Pacle J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,  
 RA Svirskas R., Taber P.B., Wan K., Stapleton M., Sutton G.G., Venter C.,  
 RA Weinstein G., Scherzer S.E., Myers E.W., Gibbs R.A., Rubin G.M.,  
 RA "Finishing a whole-genome shotgun: release 3 of the Drosophila  
 RL melanogaster euchromatic genome sequence."  
 RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426070; PubMed=12537573;  
 RA Kaminker J.S., Bergman C.M., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.E., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RA "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426069; PubMed=12537572;  
 RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
 RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.B.,  
 RA Smith C.D., Tupy J.L., Whitfield B.J., Bayraktaroglu L., Berman B.P.,  
 RA Bettencourt B.R., Celinker S.E., de Grey A.D., Drysdale R.A.,  
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,  
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
 RA Lewis S.B.,  
 RA "Annotation of the Drosophila melanogaster euchromatic genome: a  
 RT systematic review."  
 RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426071; PubMed=12537574;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.E., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RA "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426072; PubMed=12537575;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.E., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RA "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [7]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426073; PubMed=12537576;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.E., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RA "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [8]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426074; PubMed=12537577;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.E., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RA "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [9]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426075; PubMed=12537578;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.E., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RA "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [10]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426076; PubMed=12537579;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.E., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RA "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [11]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426077; PubMed=12537580;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.E., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RA "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [12]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426078; PubMed=12537581;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.E., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RA "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [13]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426079; PubMed=12537582;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.E., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RA "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [14]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426080; PubMed=12537583;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.E., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RA "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [15]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426081; PubMed=12537584;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.E., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RA "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [16]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426082; PubMed=12537585;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.E., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RA "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [17]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426083; PubMed=12537586;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.E., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RA "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [18]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426084; PubMed=12537587;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.E., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RA "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [19]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426085; PubMed=12537588;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.E., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RA "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [20]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426086; PubMed=12537589;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.E., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RA "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [21]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426087; PubMed=12537590;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.E., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RA "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [22]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426088; PubMed=12537591;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.E., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RA "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [23]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426089; PubMed=12537592;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.E., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RA "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [24]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426090; PubMed=12537593;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.E., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RA "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [25]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426091; PubMed=12537594;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.E., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RA "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [26]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426092; PubMed=12537595;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.E., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RA "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [27]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426093; PubMed=12537596;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.E., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RA "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [28]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426094; PubMed=12537597;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.E., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RA "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [29]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426095; PubMed=12537598;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.E., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RA "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [30]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426096; PubMed=12537599;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.E., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RA "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [31]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426097; PubMed=12537600;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.E., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RA "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [32]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426098; PubMed=12537601;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.E., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RA "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [33]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426099; PubMed=12537602;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.E., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RA "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [34]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426100; PubMed=12537603;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.E., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RA "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [35]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426101; PubMed=12537604;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.E., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RA "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [36]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426102; PubMed=12537605;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.E., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RA "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [37]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426103; PubMed=12537606;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.E., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RA "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [38]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426104; PubMed=12537607;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.E., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RA "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [39]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426105; PubMed=12537608;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.E., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RA "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [40]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426106; PubMed=12537609;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.E., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RA "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [41]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426107; PubMed=12537610;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.E., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RA "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [42]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426108; PubMed=12537611;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.E., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RA "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [43]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426109; PubMed=12537612;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.E., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RA "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [44]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426110; PubMed=12537613;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.E., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RA "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [45]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426111; PubMed=12537614;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.E., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RA "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [46]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426112; PubMed=12537615;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.E., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RA "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [47]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426113; PubMed=12537616;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.E., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RA "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [48]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426114; PubMed=12537617;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.E., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RA "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [49]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426115; PubMed=12537618;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.E., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RA "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [50]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426116; PubMed=12537619;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.E., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RA "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [51]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426117; PubMed=12537620;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.E., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RA "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [52]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426118; PubMed=12537621;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.E., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RA "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [53]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426119; PubMed=12537622;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.E., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RA "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [54]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426120; PubMed=12537623;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise B., Wheeler D.A., Lewis S.E., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RA "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [55]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426121; PubMed=12537624;  
 RA Celinker S.E., Wheeler D.A., Krommiller B., Carlson J., Svirskas R.,  
 RA Patel S., Frise

```

QY 30 LRELSQLEABLNKMRRLLEGSLYKPPSSSAEKVANKNDVASPLKELGLRISKFLGLD 89
DB 48 LRNPSPSEKSRNLLRNALNMGV-----PMQSHSKTMSQDLA-----DEAVILSDMFDD 98
QY 90 EEOVSQLOOCLQEDR-----GTRDSVKTUQDEROSQALILKADYYEER-----TCILRC 143
DB 99 EGFVAVELL-CTAOROOQHHPGLSRGLVAVL-----LYYDPRKAIISCTLRD 142
QY 144 VLHLL-----TYFODERHPRVAYAD-CVD-----KLEBEV----- 174
DB 143 MPQVAGSVNNTLPRGKITGLVNTYASLVDGSGILRLQLLEBNDREKCAMLTTRRA 202
QY 175 ---SKTRQO---FEEELYKTEAP---TWETHNIMTEROVSRWFVOCLEOS----- 216
DB 203 FSGKQHONQVLAGYEDIOQALAMALFNWSAQRGJ---PRHAIILMHQLANRKNHDAKGN 260
QY 217 ---MLLEIFLYAYAEEMASDILVLTAKMEKQGFSGRQTNHLDVETMDPFVDRIGYRS 273
DB 261 DVTLLIMLMAVLYAY-----DTSMLLVTEBRENHTRLPIFS 297
QY 274 ALILVEGMDIESLHKALDDBRELHOFAODGLICQMDCLM-LTFGDIPHNAVLLAMAL 332
DB 298 DEBFAE-----CFLEB-----LYAQSSMOAPRLMAIIAYSFG-----LTLAS 334
QY 333 LRHT-LNPEETS-SVVRK---IGGTAIQANVFOYLTRLLQSLASGNDCTSTACMCVYG 387
DB 335 LHHAPLOQATAIISIINRDEMILDEALGAQVFFHSLLE-----KOLVYST-QFFPR 387
QY 388 LLSFVLTSELHTLNGQODIID-----TACBVLADPS---LPELFMGTEPTSGIGILL 437
DB 388 RVHLIIT-----DFIDPMNSKVSELGRADESSHITISFLEGELEBP----- 429
QY 438 DSVCGMFPHLS--PLLQLLRALVSG-----KSTAKV--Y 469
DB 430 -----PULDSNFELMLCVAKMGDPRTVIRLCNEVWPGDPNGSTAFKXTSSVSLP 482
QY 470 SFLDKMSFNEIYKAKPHDVISHEDGTLRRROTPKLLVPGQTN-----LRIP 518
DB 483 KETSLAS-----DLRPO--TLPRK---AYLKMIISGLTRTDFSGARCAPMMLRVP 524
QY 519 QGTGQVMDLDRAYLVREWEYSYSWTL-----FYCEI-----EMLHVASTADYI- 563
DB 525 QWATG-----GKAVSMDFHT--TLGNYTSMKNDFTNIGMGEETIYTRSTPKAIT 576
QY 564 ---OHQORVKEPIDLV-HKVISDLSIADCLPITSRIYMLQRLTTVISPPYDVILAS 617
DB 577 QREAEHLAVMGIIQAVAEHDEVSR-IMICDHPNMQTPQVLLGVAQAT-----PLFLKAE 631
QY 618 CVCNCLTVLA-AANPAKV-WTDLRHTGFLPFAHAPVS-SLSQMTSABGMNAGGYMLLANS 674
DB 632 ILHTLAALAKSKETARVIMFHLBASQIIPV--PVSRSYAQCSLLEBEM--QIEBRS 684
QY 675 EOPQGEVYVTLAFLRLITLVKYGQSTOSOG-LVPCWVFVJLKMPLS---YHKMRYS 729
DB 685 EQ-----YMLTGILLOLTLMTTNPKSISGAGQRPAYEGITKXMINSVLKFTRAYKV 740
QY 730 HGVBEOICGLILELIHALINTCHETDLSHSHTPS-LOPLCICSLAYTEAGOVINIMIG 788
DB 741 PSEKQVGAQCCIKLLIYLL-----ATYRPSAMDFL----- 770
QY 789 VDTID-----WMAAQRSDGAE-----GQGGQGLIKTYKL----- 820
DB 771 -ETVDEBPYPGFHWMOLOLKSSEMLOLLCIVEBVRERLDVYNRQGGKLEBESLYALL 829
QY 821 ---AFSVTNVIRLKPPSN---VVSPLQAL-----SOHGAHNNLIVLAKRIYKH 867
DB 830 ILBAALAKONAFPEHSAANCPILLSGANMLDLNPSRKPDRHYLTI---KTV--TY 883
QY 868 DPALPRLAIQILKRLATVP-----MSVYACIGANDAAIRDAFLTRLOSKI-----E 914
DB 884 NSMPLRHALAIKILIASVTQLPNVSTQILSMYGQGSNEKLEIROGFVACLEMEVCGVGH 943
QY 915 DMRIKVMIL-----EFLTV-----AVETQGLIELF 940

```

```

DB 944 DOLLQOLANHVPLFGEDDLDNEREMSGEREYISTISQELQAGDGGLSKRPSTIELQ 1003
QY 941 L-----NLBYK-----DSDGSKFSG-----MWSCLHAVLELI----- 970
DB 1004 LKEALITKLENTLSQOLFNFVYFLFGDVLRFPFMANGRQHLAIEMQSSVNSVVLLEX 1063
QY 971 -DSQODRKYCPPLHRAAIA-----FLHLMQDRDSAMLV-----LRTX 1010
DB 1064 MKQHSBKYCE--HTARIYERITHPHGLCANRRTETILRYFLTCSDPLHNLHLSL 1120
QY 1011 P-----KFWENITSPLEGT----- 1024
DB 1121 PRQRREBDVLAHMHNLNCYSIDVKLAAGQMTFRQMDIILIGMERSHGLSWE 1180
QY 1025 ---LSPPS-----ETSPSIL-ETCALIM 1044
DB 1181 LQHSLSIQPSSSFLAMDVLPAQGSISAVAHGASASLGAFTGVKSLPILQFTSQGLH 1240
QY 1045 KIICLEIYVVKGSIDQSLKOTLKKFSEIKRFAYMSGYKSLAVH---VAFBSSCTS 1100
DB 1241 VTRLDILVLRAGTISO-----POLPEPDGHLITQLLRDCSABAGANSRAN 1288
QY 1101 LLEYQMLVSANRMLLIATTHADIMHLDTSVY---RRQLFDVLDGTRKALLVPAZYNC 1156
DB 1289 LINV-----RKLYIIT--DELMVQSIISAGQRAI-----STREMLHNAVNI 1332
QY 1157 LRLSMAKTTLLIILRQKRELSGVDELGPLTEIL-----EGVLQADQOLMEKTKARVP 1211
DB 1333 NRVETORCA-TIAPMAAM-----GQVQVLESNMEDAVILPATQRORH----- 1373
QY 1212 SAFITVLOMKEMKYSDIPOYSQVLIVANCETOEBEYILFQDTRSL--ALGSATED--K 1266
DB 1374 -IIDIVKULIKQPI---QPIIEISVQVSEVLLILANLRCCYQAEQSPDLAAR 1427
QY 1267 DSMETDCCSRHRDQBDGVCVLGLHLAKELCEVDEBDS----- 1306
DB 1428 DLSNGNGNAIGNSQVVALC-----LGRRAIGTSGDGGSGGRDGGSGNSMLRPIIK 1482
QY 1307 ---WLOVTR-----RLPIPLITLTVLSEVLSRKONLHPRATLHLLITLAPQOGATA 1356
DB 1483 NLVQMINISEVKSQKLEINIIYSSILNCLRIAKRLTBEHE----- 1523
QY 1357 VAGAGITQISICLPLLSVYQ---LSTNGTAQTPSASRSLDAPSWPGVYRLSMSIM----- 1408
DB 1524 -----YQETLISROESARTYKQGRDD-----RLRKMAARVY 1558
QY 1409 ---EOLKTLRYNLEBALDFGVGHQERTLOCLNAVTVQSLACLEBADHTVPILQLS 1464
DB 1559 GTFGBKLIIDYICH---DAVYGHQVCRMLALACIDMISELOAVSTLCDPVASRGYLGHL 1614
QY 1465 NEMKEMPHL-----POLMRDIOV---NLGYLCOACTSLHSRRKLOHLYLONKNGDL 1514
DB 1515 ESLDQSTALCGIIOVPDHLRPLUYVESRMAFLPTRANANSVGARLLL-----TEOALGV 1669
QY 1570 LSNMKVYDQPDLSKSSLENNEPQFLPSIDR-----FRSILPPLSLCDALVNS 1720
QY 1572 FTFDPVQIILDOSLDLAENFLFA-----LSFTTPTRD 1604
DB 1721 LGR-----RNSNAVQVNLFLFAHIDMVEAMRLSATPYPD 1755

```

Search completed: December 30, 2004, 15:13:50  
 Job time : 341.065 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: December 30, 2004, 15:00:06 ; Search time 55.809 Seconds

(without alignments)  
2083.096 Million cell updates/sec

Title: US-10-719-385-2

Perfect score: 9007  
Sequence: 1 MIRSKRTSVLSFGRSSREL.....PESQEPRLQVQAFVRMR 1753

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 150 summaries

Database : Issued Patents AA.\*  
1: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep.\*  
2: /cgn2\_6/ptodata/1/1aa/6A.COMB.pep.\*  
3: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep.\*  
4: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep.\*  
5: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep.\*  
6: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	216	2.4	1131	4	US-09-248-796A-19110
2	147	1.6	2954	4	US-09-150-867-1
3	146	1.6	2549	4	US-08-265-967C-1
4	146	1.6	2549	4	US-08-305-790B-2
5	144.5	1.6	2549	4	US-08-471-112A-3
6	144.5	1.6	2549	4	US-09-950-634-3
7	143	1.6	2549	4	US-09-538-0672-1112
8	143	1.6	2549	5	PCR-US95-0672-1112
9	140	1.6	3878	4	US-09-914-259-11
10	136.5	1.5	919	4	US-09-270-767-44725
11	136	1.5	2802	3	US-09-542-331-1
12	136	1.5	2802	3	US-09-510-791-1
13	131.5	1.5	147	4	US-09-270-767-33392
14	131.5	1.5	147	4	US-09-270-767-48609
15	130.5	1.4	1383	4	US-09-540-236-3580
16	129.5	1.4	1826	4	US-09-198-452A-113
17	129	1.4	3210	4	US-09-538-092-1154
18	129	1.4	3248	4	US-08-353-700-1
19	129	1.4	3248	5	PCR-US95-16216-1
20	128.5	1.4	2938	5	PCR-US94-00198-3
21	127.5	1.4	3433	4	US-09-538-092-1136
22	127	1.4	2470	4	US-08-265-967C-2
23	127	1.4	2470	4	US-08-305-790B-3
24	126.5	1.4	1106	4	US-09-248-796A-15028
25	125.5	1.4	756	3	US-09-085-199B-9
26	125.5	1.4	1064	3	US-08-726-214-8
27	125	1.4	914	3	US-09-085-199B-4

28	124.5	1.4	1824	3	US-09-228-246-2	Sequence 2, Appli
29	123.5	1.4	1090	3	US-09-085-199B-5	Sequence 5, Appli
30	123.5	1.4	1824	3	US-08-680-327-3	Sequence 3, Appli
31	123.5	1.4	2818	3	US-09-542-331-2	Sequence 2, Appli
32	123.5	1.4	2818	3	US-09-510-791-2	Sequence 2, Appli
33	123	1.4	2749	4	US-09-385-222A-4	Sequence 4, Appli
34	122	1.4	903	3	US-09-228-986-78	Sequence 78, Appli
35	122	1.4	903	4	US-10-101-464A-78	Sequence 4, Appli
36	122	1.4	1958	4	US-10-028-946-2	Sequence 2, Appli
37	122	1.4	2054	4	US-10-028-946-2	Sequence 2, Appli
38	122	1.4	3066	3	US-08-952-127-12	Sequence 12, Appli
39	121	1.3	2482	1	US-08-328-254-6	Sequence 6, Appli
40	120.5	1.3	2987	2	US-08-970-269A-29	Sequence 29, Appli
41	120.5	1.3	2987	3	US-09-407-562-29	Sequence 29, Appli
42	120.5	1.3	3672	2	US-08-822-445-12	Sequence 12, Appli
43	120.5	1.3	3672	2	US-09-336-540-12	Sequence 12, Appli
44	120.5	1.3	3801	2	US-08-822-445-10	Sequence 10, Appli
45	120.5	1.3	3801	2	US-09-336-540-10	Sequence 10, Appli
46	120.5	1.3	3959	2	US-08-970-269A-30	Sequence 30, Appli
47	120.5	1.3	3959	3	US-09-407-562-30	Sequence 30, Appli
48	119.5	1.3	2930	3	US-09-417-822-2	Sequence 2, Appli
49	119	1.3	2713	5	PCR-US96-01735-1	Sequence 1, Appli
50	119	1.3	3119	1	US-08-246-982A-16	Sequence 16, Appli
51	119	1.3	3119	1	US-08-453-265-16	Sequence 16, Appli
52	117.5	1.3	2020	1	US-07-551-531-2	Sequence 2, Appli
53	117.5	1.3	2818	1	US-08-510-284-1	Sequence 1, Appli
54	117.5	1.3	2818	1	US-08-411-389-2	Sequence 2, Appli
55	117.5	1.3	2818	2	US-08-449-933-2	Sequence 2, Appli
56	117.5	1.3	2818	3	US-07-966-049A-2	Sequence 2, Appli
57	116.5	1.3	1251	4	US-09-252-991A-17263	Sequence 17263, A
58	116.5	1.3	2871	4	US-09-538-092-336	Sequence 936, App
59	116.5	1.3	3433	4	US-09-091-501B-10	Sequence 10, Appli
60	116	1.3	585	1	US-09-656-479A-20	Sequence 20, Appli
61	116	1.3	1190	4	US-08-337-650A-2	Sequence 2, Appli
62	116	1.3	1190	3	US-09-048-887-2	Sequence 2, Appli
63	116	1.3	2485	5	PCR-US94-00198-1	Sequence 1, Appli
64	116	1.3	2485	5	PCR-US94-00198-2	Sequence 2, Appli
65	115.5	1.3	1179	4	US-09-792-024-89	Sequence 89, Appli
66	115.5	1.3	1441	4	US-09-252-991A-28143	Sequence 28143, A
67	115.5	1.3	3056	4	US-09-360-416-2	Sequence 2, Appli
68	115	1.3	1110	4	US-09-270-767-42166	Sequence 42166, A
69	114.5	1.3	1077	4	US-09-412-210-1	Sequence 1, Appli
70	114.5	1.3	1077	4	US-10-121-911A-1	Sequence 1, Appli
71	114.5	1.3	2186	2	US-08-822-445-2	Sequence 2, Appli
72	114.5	1.3	2186	3	US-09-396-540-2	Sequence 2, Appli
73	114	1.3	1299	3	US-08-460-900C-62	Sequence 62, Appli
74	114	1.3	1299	3	US-08-674-509B-48	Sequence 48, Appli
75	114	1.3	1299	3	US-08-954-698-48	Sequence 48, Appli
76	114	1.3	1299	4	US-09-639-695-62	Sequence 62, Appli
77	114	1.3	1299	4	US-09-448-188-48	Sequence 48, Appli
78	114	1.3	1299	4	US-08-954-188-48	Sequence 48, Appli
79	114	1.3	1299	4	US-08-954-740-48	Sequence 48, Appli
80	114	1.3	1487	4	US-09-489-039A-12113	Sequence 12113, A
81	114	1.3	2220	4	US-09-335-011-1	Sequence 1, Appli
82	114	1.3	3051	3	US-08-934-386-9	Sequence 9, Appli
83	113.5	1.3	3051	4	US-09-360-416-3	Sequence 3, Appli
84	112.5	1.2	572	4	US-09-543-681A-8138	Sequence 8138, Ap
85	112.5	1.2	949	4	US-09-538-092-1290	Sequence 1290, Ap
86	112.5	1.2	1403	4	US-09-705-872-1	Sequence 1, Appli
87	112.5	1.2	2108	4	US-09-538-092-87	Sequence 87, Appli
88	112	1.2	15281	2	US-08-471-119A-2	Sequence 2, Appli
89	111.5	1.2	1286	3	US-09-268-140-3	Sequence 3, Appli
90	111.5	1.2	2349	4	US-09-538-092-914	Sequence 914, App
91	111.5	1.2	2644	4	US-09-028-047C-2	Sequence 2, Appli
92	111.5	1.2	3056	4	US-08-508-836A-8	Sequence 8, Appli
93	111.5	1.2	3056	2	US-08-629-001A-3	Sequence 3, Appli
94	111.5	1.2	3056	2	US-08-874-266-2	Sequence 2, Appli
95	111.5	1.2	3056	3	US-08-642-274D-3	Sequence 3, Appli
96	111.5	1.2	3056	3	US-08-952-127-3	Sequence 3, Appli
97	111.5	1.2	3056	3	US-08-952-014C-3	Sequence 3, Appli
98	111.5	1.2	3056	4	US-08-984-090-2	Sequence 2, Appli
99	111	1.2	862	4	US-09-665-479A-14	Sequence 14, Appli
100	111	1.2	1206	4	US-09-245-928A-19	Sequence 19, Appli

101	111	1.2	1243	4	US-09-245-928A-17	Sequence 17, Appl
102	111	1.2	1247	1 <td>US-08-472-93A-10</td> <td>Sequence 10, Appl</td>	US-08-472-93A-10	Sequence 10, Appl
103	111	1.2	1247	2 <td>US-08-323-460A-10</td> <td>Sequence 10, Appl</td>	US-08-323-460A-10	Sequence 10, Appl
104	111	1.2	1247	2 <td>US-08-461-146C-10</td> <td>Sequence 10, Appl</td>	US-08-461-146C-10	Sequence 10, Appl
105	111	1.2	1257	3 <td>US-08-461-145C-10</td> <td>Sequence 10, Appl</td>	US-08-461-145C-10	Sequence 10, Appl
106	111	1.2	1257	3 <td>US-08-947-823-5</td> <td>Sequence 5, Appl</td>	US-08-947-823-5	Sequence 5, Appl
107	111	1.2	1257	3 <td>US-09-245-928A-18</td> <td>Sequence 18, Appl</td>	US-09-245-928A-18	Sequence 18, Appl
108	111	1.2	1597	3 <td>US-09-423-890-13</td> <td>Sequence 13, Appl</td>	US-09-423-890-13	Sequence 13, Appl
109	111	1.2	1597	3 <td>US-08-628-823-14</td> <td>Sequence 14, Appl</td>	US-08-628-823-14	Sequence 14, Appl
110	111	1.2	2008	4 <td>US-09-270-767-43754</td> <td>Sequence 43754, A</td>	US-09-270-767-43754	Sequence 43754, A
111	111	1.2	2008	4 <td>US-09-270-767-46774</td> <td>Sequence 46774, A</td>	US-09-270-767-46774	Sequence 46774, A
112	111	1.2	804	4 <td>US-09-270-767-46750</td> <td>Sequence 46750, A</td>	US-09-270-767-46750	Sequence 46750, A
113	110.5	1.2	2261	4 <td>US-09-526-193A-1</td> <td>Sequence 1, Appl</td>	US-09-526-193A-1	Sequence 1, Appl
114	110.5	1.2	3144	1 <td>US-08-246-982A-6</td> <td>Sequence 6, Appl</td>	US-08-246-982A-6	Sequence 6, Appl
115	110.5	1.2	3144	1 <td>US-08-453-265-6</td> <td>Sequence 6, Appl</td>	US-08-453-265-6	Sequence 6, Appl
116	110.5	1.2	3144	2 <td>US-08-457-273B-42</td> <td>Sequence 42, Appl</td>	US-08-457-273B-42	Sequence 42, Appl
117	110.5	1.2	3144	3 <td>US-08-556-418-21</td> <td>Sequence 21, Appl</td>	US-08-556-418-21	Sequence 21, Appl
118	110.5	1.2	3144	3 <td>US-09-041-886-15</td> <td>Sequence 15, Appl</td>	US-09-041-886-15	Sequence 15, Appl
119	110.5	1.2	3144	4 <td>US-09-538-092-1118</td> <td>Sequence 1118, Ap</td>	US-09-538-092-1118	Sequence 1118, Ap
120	110	1.2	1575	4 <td>US-09-917-254-83</td> <td>Sequence 83, Appl</td>	US-09-917-254-83	Sequence 83, Appl
121	110	1.2	1867	4 <td>US-09-824-574-5</td> <td>Sequence 5, Appl</td>	US-09-824-574-5	Sequence 5, Appl
122	109.5	1.2	1055	3 <td>US-09-031-563-27</td> <td>Sequence 27, Appl</td>	US-09-031-563-27	Sequence 27, Appl
123	109.5	1.2	1055	4 <td>US-09-392-277-27</td> <td>Sequence 27, Appl</td>	US-09-392-277-27	Sequence 27, Appl
124	109.5	1.2	1055	4 <td>US-09-258-000-27</td> <td>Sequence 27, Appl</td>	US-09-258-000-27	Sequence 27, Appl
125	109.5	1.2	1315	3 <td>US-09-031-563-2</td> <td>Sequence 2, Appl</td>	US-09-031-563-2	Sequence 2, Appl
126	109.5	1.2	1315	3 <td>US-09-031-563-25</td> <td>Sequence 25, Appl</td>	US-09-031-563-25	Sequence 25, Appl
127	109.5	1.2	1315	3 <td>US-09-293-505-10</td> <td>Sequence 10, Appl</td>	US-09-293-505-10	Sequence 10, Appl
128	109.5	1.2	1315	4 <td>US-09-392-277-2</td> <td>Sequence 2, Appl</td>	US-09-392-277-2	Sequence 2, Appl
129	109.5	1.2	1315	4 <td>US-09-392-277-25</td> <td>Sequence 25, Appl</td>	US-09-392-277-25	Sequence 25, Appl
130	109.5	1.2	1315	4 <td>US-09-258-000-2</td> <td>Sequence 2, Appl</td>	US-09-258-000-2	Sequence 2, Appl
131	109.5	1.2	1315	4 <td>US-09-258-000-25</td> <td>Sequence 25, Appl</td>	US-09-258-000-25	Sequence 25, Appl
132	109.5	1.2	1315	4 <td>US-09-060-939A-10</td> <td>Sequence 10, Appl</td>	US-09-060-939A-10	Sequence 10, Appl
133	109.5	1.2	2662	4 <td>US-09-559-684B-31</td> <td>Sequence 31, Appl</td>	US-09-559-684B-31	Sequence 31, Appl
134	109	1.2	553	4 <td>US-09-046-572-3</td> <td>Sequence 3, Appl</td>	US-09-046-572-3	Sequence 3, Appl
135	109	1.2	2663	4 <td>US-09-538-092-1252</td> <td>Sequence 1252, Ap</td>	US-09-538-092-1252	Sequence 1252, Ap
136	108.5	1.2	1087	4 <td>US-09-914-259-12</td> <td>Sequence 12, Appl</td>	US-09-914-259-12	Sequence 12, Appl
137	108.5	1.2	1186	4 <td>US-09-543-681A-5514</td> <td>Sequence 5514, Ap</td>	US-09-543-681A-5514	Sequence 5514, Ap
138	108.5	1.2	1512	4 <td>US-09-328-352-5163</td> <td>Sequence 5163, Ap</td>	US-09-328-352-5163	Sequence 5163, Ap
139	108.5	1.2	3118	2 <td>US-08-457-273B-8</td> <td>Sequence 8, Appl</td>	US-08-457-273B-8	Sequence 8, Appl
140	108	1.2	778	5 <td>PCT-US93-03076-3</td> <td>Sequence 3, Appl</td>	PCT-US93-03076-3	Sequence 3, Appl
141	108	1.2	1513	5 <td>PCT-US93-03076-2</td> <td>Sequence 2, Appl</td>	PCT-US93-03076-2	Sequence 2, Appl
142	107.5	1.2	2482	4 <td>US-09-252-991A-16967</td> <td>Sequence 16967, A</td>	US-09-252-991A-16967	Sequence 16967, A
143	107	1.2	1572	4 <td>US-09-710-279-2906</td> <td>Sequence 2906, Ap</td>	US-09-710-279-2906	Sequence 2906, Ap
144	106.5	1.2	335	2 <td>US-08-875-062-3</td> <td>Sequence 3, Appl</td>	US-08-875-062-3	Sequence 3, Appl
145	106.5	1.2	1232	4 <td>US-09-248-796A-14563</td> <td>Sequence 14563, A</td>	US-09-248-796A-14563	Sequence 14563, A
146	106.5	1.2	3174	2 <td>US-08-477-451-3</td> <td>Sequence 3, Appl</td>	US-08-477-451-3	Sequence 3, Appl
147	106	1.2	1051	4 <td>US-09-543-681A-8143</td> <td>Sequence 8143, Ap</td>	US-09-543-681A-8143	Sequence 8143, Ap
148	106	1.2	1068	4 <td>US-09-248-796A-16119</td> <td>Sequence 16119, A</td>	US-09-248-796A-16119	Sequence 16119, A
149	106	1.2	1530	4 <td>US-09-976-594-736</td> <td>Sequence 736, Ap</td>	US-09-976-594-736	Sequence 736, Ap
150	106	1.2	1757	4 <td>US-09-724-126A-15</td> <td>Sequence 15, Appl</td>	US-09-724-126A-15	Sequence 15, Appl

## ALIGNMENTS

RESULT 1  
 US-09-248-796A-19110  
 ; Sequence 19110, Application US/09248796A  
 ; Patent No. 6747137  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Ketch Weststock et al  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN  
 ; FILE REFERENCE: 107196.132  
 ; CURRENT APPLICATION NUMBER: US/09/248, 796A  
 ; CURRENT FILING DATE: 1999-02-12  
 ; PRIOR APPLICATION NUMBER: US 60/074, 725  
 ; PRIOR FILING DATE: 1998-02-13  
 ; PRIOR APPLICATION NUMBER: US 60/096, 409  
 ; PRIOR FILING DATE: 1998-08-13  
 ; NUMBER OF SEQ ID NOS: 28208  
 ; SEQ ID NO 19110  
 ; LENGTH: 1131

```

; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-19110
Query Match
Best Local Similarity 18.5%; Pred. No. 2,4e-11;
Matches 222; Conservative 230; Mismatches 441; Indels 308; Gaps 60;

2.4%; Score 216; DB 4; Length 1131;

19 ELWTLLGRSALRE-----LSQIEALEINKMRRLLEGLSYKPPSSAER 64
52 QYWFEDNALSLRNCQDPYALADLEFLLSKEVTLN-----PSPFLTKN 96
65 YKANQDVASPLKEIGLR--ISKELGDEBSYQLOCYLOEDRGKTSYKTVQDR-- 120
97 DKQSPNIQT--RDISLKGIIITNYSVDISDGKLSNMLND--VLETRVICQTNKKI 151
121 -----QSQALILIKIAD-YYE-----ERTCILRCVHLITFYODERAPYR 159
152 PCKTAPQLEAIKSKLHKDKKYENKRLRLYSKILRRRIIUKITVELL---NKSNSYA 208
160 VEVADCVDKLEKELVSKYRQFEBL-----YKTEAPYETHGNIMTERQV-SRWVFQ 210
209 ---SSSIQNGKEIFLS-NQYLESLIVYIGNASHIMKSSYITGINKIEIDETINETVLF 264
211 CLREOSMLRITFLYYAFYFEMAPSDLLVTKMFEQGFSGRQTNRHVDEBTMPFVDRIG 270
265 CIEKCKLIEL-----SVQNAVDAQAVHLMKRM---RDTNYSV---ALGPYV---S 308
271 YFSALILVEGMDI-----ESLHKCALDRREIHOFPADGL----- 305
309 YHBAFSLIQLEFVTLTIQYLNNSPFSVETLSSYMD-VQVFYKVDHAIANPANNIF 367
306 -IQGDMDCML-----TFGDIPIHAPVILNAL--LRHTNPEETSSVVRKIGTAIQ 355
368 VVYSGSIIILRRFYLOEYPELNSKEFLSQFNLSQLKRTIN-----LVNOKCDN 418
356 LNVFOYLTRLRLOSIAAGNDCTTSTACMCVYG--LISFVLTSLRLTGNQOIIDPAC 413
419 LDFSSILKUNELKNDK-----LYSAILTLIIASIPLTITSEV---TSCI 463
414 VLADPSLP---ELFNGTEPTSGIILDSVCGMPFHLSPILQILRALVSGSKAKVY 469
464 LSVIGCPNNVVISFFENNATQNAIIART---KFLPILISPYIQV--ASINGNFA---LH 515
470 SFLDKSPFVELYKHPHDIVISHEGTLWRORPKLL--YP---LGGQTNLRIPQSTVQ 524
516 EFPDLKSYIOVFKEKERNNNYQIDQNTBLVKTFTIDYVPRPANKKLSVLSIGTAK 575
525 VM--LDDRATLVNWEYSYSWTLPFCIEIMLHVSTADVTIOHQCRVKPIIDLVRKV--I 580
576 ILPSANPDEVLTFLYNNYGMALGRVLQNVSKIFNNSD---SETMELVINILNSLNNV 631
581 STDLSTADCLIPR-----TSRIYMLQRLTT--VISPPVDVIVASCNCLTVLAARNP 630
632 VIDNGVDSKVMLEANSAYTDSIDILEVILRLLEQGLHRLNRYVLVSVNLTNLNPFIS 691
631 AKVWTDLRHTGFLPFAHPPVSSISQMTISABGMNAGGYMLNMSQPOGEYGTIAFLR 690
692 YRW-----PYLSK--SALFAQNGKEGLAAVIFGSI-----EMVNGDYNTVSLIKL 736
737 AEAALIONCLSDQDYPEKESYVI-MLRFGHVLDPETFLCRYNBOCQKLEIGVLLDT 795
744 IHAIIULCHETD-----LHSHTPSLQFCIGSIATYAG 778
796 FSTILLASVIGDECVANTKTVKVPADAASRLIDSEFLSDEDSPPAARLIT----- 846
779 QTVINIMIGIVDTIDVMAAOPRSDGABEGOGQLIKTVKLAFSVYNNVRLKPSNIV 838
847 -TMIEHLQSDLDLIELTSSSFMYD-----NMHICALSSRLIITIR-TSEHL 892
839 SPL--EQALSQGHGNNLIIVLAKYIYHKHDPALPRLAIQLKRLAT-----VAPMSV 890

```

Db 893 KPLAEKSLF---TKTSDLVLSRPFESVRKD-----ILDLLTALTSQDMPDGTAP--SL 942  
Qy 881 YACLGND-AAAIRDA-FRLRLQSKIEDMRKIMLEFLVAVV--TOPGLIEFL----- 941  
Db 943 LSHLENAQVILHSLADLDNSPDDYKRIKSLYDFICAVMEGKQEGLAIVLFTGRDVG 1002  
Qy 942 NLEVDGSDGSKESFSGMWSCLHVLIELIDSGQDRWYCP--PLHRA-AIAFLHALMD 998  
Db 1003 DYSADTDEATKSL-----LQILKKIRIMRYENSVSHLYDALALACNSMTT 1053  
Qy 999 RDSAMLVLRTPKPFEMNLTSPFLSPSETSPSI-----LETCALIMKIICLEY 1053  
Db 1054 AKES-----EHDDEFIQTLIGRVKQILDPPDSSSFISRCVELKLVSKIAELIILYFT 1108  
Qy 1054 V 1054  
Db 1109 V 1109

## RESULT 2

US-09-150-867-1  
Sequence 1, Application US/09150867  
Patent No. 6645748  
GENERAL INFORMATION:  
APPLICANT: Wood, Kenneth W.  
APPLICANT: Sakowicz, Roman  
APPLICANT: Goldstein, Lawrence S.B.  
APPLICANT: Cleveland, Don W.  
TITLE OF INVENTION: The Regents of the University of California  
TITLE OF INVENTION: Plus End-directed Microtubule Motor Required for  
FILE REFERENCE: 18557C-000110US  
CURRENT APPLICATION NUMBER: US/09/150,867  
CURRENT FILING DATE: 1998-09-10  
EARLIER APPLICATION NUMBER: US 60/058,645  
EARLIER FILING DATE: 1997-09-11  
NUMBER OF SEQ ID NOS: 11  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 1  
LENGTH: 2954  
TYPE: PRT  
ORGANISM: Xenopus sp.  
FEATURE:  
OTHER INFORMATION: Xenopus centromere-associated protein-E (XCEP-E)  
OTHER INFORMATION: member of the kinesin superfamily of microtubule  
OTHER INFORMATION: motor proteins  
FEATURE:  
NAME/KEY: DOMAIN  
LOCATION: (1)..(472)  
OTHER INFORMATION: kinesin like motor domain  
FEATURE:  
NAME/KEY: DOMAIN  
LOCATION: (473)..(2752)  
OTHER INFORMATION: rod domain  
FEATURE:  
NAME/KEY: DOMAIN  
LOCATION: (2753)..(2954)  
OTHER INFORMATION: tail domain  
US-09-150-867-1

Query Match 1.6%; Score 147; DB 4; Length 2954;

Best Local Similarity 17.1%; Pred. No. 0.0012; Indels 508; Gaps 57;  
Matches 244; Conservative 202; Mismatches 474;

Qy 557 VSTADVIOHCORVKPII-----DLVHKVISTDLSTADCLLP-----TSRI 597  
Db 1 MSEGAVAVCVRRRLIQREGDQANLQKAGNNTISQVDGKSNFDRVFNHSHSTQI 60  
Qy 558 YMLLRLTTVTSPPVDVASCV---NCLTVLAANPAKVTDLRHGTGLFPVAHPVSL 653  
Db 61 YOEI-----AVPIIRSLAQYNGTIFAYGOTSQKTYMTMGTPNSLIGIIPAIOBV 111

Qy 654 SOMISAEGMAGYGNILMNSQPGEGYVIAFLRLITLVKQL----- 699  
Db 112 FKII-----GEIPREFLAVSYNEIYNEVYKOLLCDRRKKPLEIRED 155  
Qy 700 -----GSTQSGVLP--CVMFLK--EMLPSYHKRMYNSHGVREQIGCLILEIHAIL 748  
Db 156 FRNRYVADLTIELVAVBEHYIOWIKKEGRKHVGETKNDHSSRS-----HTIF 205  
Qy 749 NLCHETDLSHSTPBLQCLTCSLAYTAGQVINKIGVDTIDMVAAPRSDAGQ 808  
Db 206 RPIYESRNDPTNSEN-----CDGA-----VMSHNLVDLASERSAQGAAG-- 250  
Qy 809 GGGOLLITVLAESVTNNVRLKPSN-----VSPLEQALSGHAGNNLIJAVLAKY 862  
Db 251 -----VRLKEGCNINSFLIGQVIKKLSQAGG-----F 281  
Qy 863 IYHKDPALPRLAIQLKR-----LATVAPMSVYACI-----GNDAAIR----- 902  
Db 282 INYR--DSKLTRILQNSLGNAKTVIICITIVPSFDETISTIQFAPSTAKHVNTPHVNEVL 340  
Qy 903 --DAFLRLQSKIEDMRKIMLEFLVAVETOPGLIELFLNLEVKDSDGSKESLGMW 960  
Db 341 DDEALLKRYRKEILDLKKQLENLE-----SSSETKAQAMAKEERTQ--- 381  
Qy 961 SCLHVLIELIDSGQDRWYCPPLHRAAIAFLHALMDRDSAMLVLRTPKPFME----- 1015  
Db 382 --LAEITQLKKEHEDRW---HLTNIVVASSQESQQR-----VKKRYTMAPKIQ 431  
Qy 1016 -----NLSPFLGTLSPSETSE--PSILETCALIMKIICLE----- 1050  
Db 432 NSLHAGVSDPDMLSRLPGNFSKKAKFSDMSPFE-----IDSVCTEFSDFDALSMDS 487  
Qy 1051 -----IYVVGSDLSQSLD-----TLK 1068  
Db 488 NGIDAEMNLASKYTHREKTSLSQSMIDGQISDSVQFHSSKENQOYLPRDQGMABCR 547  
Qy 1069 KFSIEKPAVW-----SGVKSIAHVATEGSSCTSLLEYQMLVSAMRMLLI 1117  
Db 548 KASFEKEITSLQOQLOSKEBEKELVQSFELKIALEBQLSVKAKNLEMVNS--REHIN 606  
Qy 1118 ATTHADIMHLDTSVVRQLFLDLDTGTRALL--LVPAVNCRLRGS--KCTLLILLRLQ 1173  
Db 607 AEVQTDV--EKEVVRKEMSVLGDGYNASNDLQDSSVDGKRSLSSHDECIERKMLEQ 663  
Qy 1174 WKRELSYDELIGPITELLEGVLQADQ-----QLMKTKAKYFSAFIVLQKEMKV 1225  
Db 664 --KIVDLEEFIEMLNKKSENDKOKSSBQDMESIQLCEALMAEKANALBEALMRD--- 717  
Qy 1226 SDIPQSOVLNVGCTLOEVIALFPDQTRHSIALGATEDKSDMETDCCSRSHRDQRDG 1285  
Db 718 ----NFDNIILIE--NETLKRBIADL-----ERSL-----KENQETNEFELKEKTOKEH 760  
Qy 1286 VCVIGLHLAKELCEVEDDEGSMWQ--VTRRLPILPTLLTLEVSL-----RMKONLFTBAT 1340  
Db 761 EAQL--IHETIGSLKKLVENAMYNQVLBEDLFTKTKLKEGHIQLAELKRADNLOKVRN 819  
Qy 1341 LHLILTLARTOOGATAVAGAGITQSICLPILSVYQLSTNGTAQTPSAR----- 1389  
Db 820 FDLVSVM-----GDSEGLCEIIFQLKQSLSDADAVTRDAQKECSFLSENILE 866  
Qy 1390 ---KSLDAPSWPVGVRRLSMSLMEQLLKTLYRNF----- 1419  
Db 867 LKEKHEEDTSNMYNQEKRAASLFEKQLETKSNYKMEADLOKELOSANEINYLGLLAG 926  
Qy 1420 -----LPEALDFVGHQERTIQLCINAVRTV--QSLACT----- 1450  
Db 927 KVPRLDLGRVLEKKEVSEFSKQLEKALEEKNA--LENEVTCLSEYKFLPNEVECLNQISK 985  
Qy 1451 -----EADHTVGPITLQLSNFMKEMHHLQPLMDIQVNLGYLQACTSLLSHKM 1501  
Db 986 ASEBIMLLKQEGHSSASIIISQEIIMQOSQOILDTDEVHTTOSKVOQTEBOYLEMKM 1045  
Qy 1502 ----LQHYLQNK-----NODGLPSAVAGRVQRPAAASAPSSKQPAADTEAS 1546





```

Db      943 JDEFPYAVSMV-----ALMRFIFROGSIHHHTTWVOAITFISLGLKCYQFLPQVA-- 994
Oy      1109 SAMRMILLIATTHADIMHILTDSVVRRLFDLV-----LDGTALLIVPASV 1155
Db      995 -----PRFLNIVRCDAIREFIFOQLGMVLSEVFKSHIREPYMEDEIYLNEFEFWM 1044#
Oy      1155 NCLRLGSMKCFTLLILLRQMKRELGSVEIIGRLPEILIEGVLOAQDOLMEKTKAVPSAF 1214#
Db      1045 NT-----SUGSTIIIL-LIEQIYVALGGERKLY--LEQLIPMLRV--FWHDNSQGRIVS- 1093#
Oy      1215 ITVIQMKEMKVSDIPQYSQVLNVNCETILOEVIYALFDQTRSL-----ALGATEDKDQM 1268#
Db      1094 IKTLAAILFGANILDVDYHLHL-----PIVKLFDAEPVLPFRKAALFTVDELTESTL 1146#
Oy      1270 E-TDDCSRHRHQDGVCVIGLHAKEICEVEDEDCSWL--QVTRLRPILTLLTTLEV 1326#
Db      1147 DFTDYASIIH-----PIVRTLDOSPBLRSTAMDTLSLVFOQKRKYQIFTPWNRKVLV 1200#
Oy      1327 SLRMKNLIH-----FT-----EATLHLLLTARTOOAGTVAGAIGTOSICPLT 1371#
Db      1201 RHRINHGXYDVLCIRIVKGYTLADEEDPLIYQHNMILASSGDALASAPVBGT----PMK 1256#
Oy      1372 SVYQSLNTGTQTQTPBASKSLDABPW-PGVYRLSNLSM-EQLIKTLR-----YNFLP 1421#
Db      1257 KLHVSTINLQKAWGAARRSKD--DMLEMRRLSELHKDSSPBRLSCWMAQAQYN--P 1312#
Oy      1422 EALD-----FYGV-----HOERTLOCINAVRTQSACLEADHVTGFILQTSNPKE 1469#
Db      1313 MARDLFNAAFVSCMSNELNQDDELIRSIELALTSQDIA---EYQT---LINLAEFM-E 1365#
Oy      1470 WHFHLPQLMRDIQ-VNLG---YLCQACTSLHSRKMLQHYLQNKNGDGLPSAVA--QRV 1522#
Db      1366 HSDKGPLELRDNGVILGRERAAKCRAYAKALHYKEL--EFQKGPPTAILESLISNNKL 1423#
Oy      1523 QRPPSAGSAARSSKQ-PAADTEASEQOALH 1552
Db      1424 QQPERAASGVLEYAMKHFGSELETIOATWYEKXH 1454

RESULT 5
US-08-471-112A-3
Sequence 3, Application US/08471112A
Patent No. 6113264
GENERAL INFORMATION:
APPLICANT: Molnar-Kimber, Katherine L.
APPLICANT: Failii, Amedeo F.
APPLICANT: Caggiano, Thomas J.
APPLICANT: Nakaniishi, Koji
APPLICANT: Chen, Yanguu
TITLE OF INVENTION: EFFECTOR PROTEINS OF RAPAMYCIN
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
ADDRESSSEE: Dunner, L.L.P.
STREET: 1300 I Street, N.W.
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471,112A
FILING DATE: 06-JUN-1995
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/384,524
FILING DATE: 13-FEB-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/312,023

```

```

1 FILING DATE: 26-SEP-1995
2
3 PRIOR APPLICATION DATA:
4
5 APPLICATION NUMBER: US 08/207,975
6
7 FILING DATE: 08-MAR-1994
8
9 ATTORNEY/AGENT INFORMATION:
10 NAME: Siekman, Michael T.
11 REGISTRATION NUMBER: 36,276
12 REFERENCE/DOCKET NUMBER: 01142.0058-00000
13
14 TELECOMMUNICATION INFORMATION:
15
16 TELEPHONE: 202-408-4000
17
18 TELEFAX: 202-408-4400
19
20 INFORMATION FOR SEQ ID NO: 3:
21
22 SEQUENCE CHARACTERISTICS:
23
24 LENGTH: 2549 amino acids
25
26 TYPE: amino acid
27
28 STRANDEDNESS: single
29
30 TOPOLOGY: linear
31
32 MOLECULE TYPE: peptide
33
34 US-08-471-112A-3

```

Query Match	1.6*	Score 144.5;	DB 3;	Length 2549;
Best Local Similarity	18.6%;	Pred. No. 0.0017;		
Matches 336;	Conservative 238;	Mismatches 628;	Indels 605;	Gaps 87
QY	237	LVLTFRMEKQFGSGRQTNRLHVLDETMTPF-----VDRIGEFSALTLVEGMDIE	284	
Db	217	LILTTQREPKMKQRPQWYRHHTFFEAKEGFDDTLAKEKMNDRRI-HGALLILNLVYRIS	275	
QY	285	SLHKCALDDRRRLHQFADGII---CODMCLMTFGDIPHHAVLLIAMALLRHTLNP	340	
Db	276	SMEEGRLL-REEMEEITQOOLVHDKYCD---LMGFQTKRHHITPTPSF---QAVQPO	325	
QY	341	ETSSVVRKIG-----GTA-----	353	
Db	326	QSNALVGLLGSSHQGLMGFGTSSPAKSTLVESRCCRDIMEKFPDQCVWLKCRNSK	385	
QY	354	--IOLANFOYLTRL-----LQSLASGNDCTTS---TACMCVYLLSFYLT	394	
Db	386	SLIGMTLNLPLRLAARPSAFTQYVLQDTNMNALSVCKEKERTAFQALGLLSAVR	445	
QY	395	SLLEHTLGNODIIDTAC-----EVLADPSLPELFWGTEPTSGAIIID-	438	
Db	446	SEFKVYLPRLVLDITIRALPRQDFAKRQKAMQVDAVFTCISMLARAMGPIQDQIKEL	505	
QY	439	---SVCCMFHLSSPLLQLRLALVSGSTAKKVSFLDKMSFYNELYHGKRRHVDVSHEDG	495	
Db	506	EPMLAVELSPALVLYDLRSQIFOLK--KDIDGGLKM--LSLVLMHKP-----	551	
QY	496	TLMRQTPK-LLYELGSGQTNLRIPQGT-VQGVMLDDBAYLVRMESYSWTLF-----	546	
Db	552	-LRRPMPKGLAHQLASPGULTLEPADVGSITLALKT-LGSFPEEGSLQFVRHCADH	609	
QY	547	-----TCE--TEMLLAVSTADVIQHCQRYKPIIDLVHKVI-----STD	584	
Db	610	FLNSEHKEIMEAARTCSRLTSPSIHLSIGHAVVSOATAQVADVLSKLVLVGITDDP	669	
QY	585	SINDCLPIRSRYML-----LQRLTVISPPVDVYASVNCULTV--LAARNA-----	631	
Db	670	DIRYCVLASLDERDAHLAQAENLQALFVALNDVFEIRELATI-TVGRSSSNMDFVMP	728	
QY	632	-----KWVTDLRHTGF-----LPFVAHPVSSLSQMTSAEGMNAAGGYN-----L	670	
Db	729	FLRKMQLQIITELHESGIRIKQGSAMGLHIVSNARLLRP-----YHEPLIKALIL	781	
QY	671	LMNSEQPGEGYGTIAFLRLITTLVKQGLSGTOSQGLVPCVMFLKEMLPYHKRRYNSH	730	
Db	782	KLQDPDDPNPVGVIYNNVIATL---GEL--AQVSGL-----EM---RKM-----	816	
QY	731	GVAREQGLLDELHIALINCHETDLSHSHTPSLQFL--CICSLAY-----	774	
Db	817	-----VDELFIIMDLQSSLSLAKQVALMTLQGLVASGYYVPEYRKYPTLLEV	867	
QY	775	-----TEAGQ-----TVYINING-----IGVTDIPVMAAPRSDAEGQGGOL	813	



```

Db      868 LLNFKTEQNGCTREAIRVIGLGLADLPYKHKVNIIGIDSRDASANSLSKSSQDSS 927
Qy      814 LIKTVKLAFTVNTVLT-RLKPPSNVSVPL-----EQALSQGHAGNNTLAVLAKYIYHK- 866
Db      928 DYSTSEMILVNNGNPLDDEFYPAVSVMALMIFRDSLSHHHTVMVQAITFFKSGIKCV 987
Qy      867 -----HDPALPRLAIOQLKRLATAPMSVYACLGNDAAARDAFL--TR 908
Db      988 QPLPQVMTPLNVIRVCGAIRBFLFOQLGMLVSVKSHIRPYMDEIVTLTKREFWMTS 1047
Qy      909 LOSKIEDMRIRKVMILEFTVAVETQ-----PGLLEFLNLEVKDSDGS-----K 953
Db      1048 IQSTI-----ILILEQIVVALGGEFKYLPQLIPHMLRVFMHDSPPRIYSIKLAAIQ 1101
Qy      954 EFSIGMWSCLHVL---ELIDSQQQDRYKCPPLHRAPIAFALHMO--DRDSAMVL 1007
Db      1102 LFGANLDDYHLHLLPPIVKLFDAPF-----APLPRKALBTVRLTESLDFDYASRII 1156
Qy      1008 RTKRFKFWENLSPLEGTSPSPSEISLFTCALIMK----- 1045
Db      1157 H-----FYRITLDQSPBLRSTIAMDTLSSLVFLQKKYQIFIPMVKLVRRIN 1205
Qy      1046 -----IICLEIYVYVKG-SLDSQSLKDTL----- 1067
Db      1206 HQRYDVILC-----RIVKGYTLADEEDPLIYQHRMLRSGGDALASGVENTGPMKXUHS 1261
Qy      1068 -----KFSIEKRFAY--WSGVKSLAVHVBETGS---SCTSLLE-YQML----- 1107
Db      1262 TINLQKAMGARARVSKDMLRMLRSLLELKDSPPSLRSCWMLAQAYNMAPDLFNAA 1321
Qy      1108 -VSAMRML-----LIATHADIMHMLTDSVVRQQLFDVLDCTKALLVPSVN 1155
Db      1322 FVSCMSELNEDQOELRSIELATLSODIAETQTLMLAEFMESDGPPLRDNGIV 1381
Qy      1156 CLRLSGMKCTLLILLRQMKBELGSVDILGLPTEILRGVLAQOQLMEKTKAVFSAFI 1215
Db      1382 LIGRAACRAVAAKALHKELEPKFCK-----GPTAILESLISINKKQPPRA---AGV 1432
Qy      1216 TVLQMK---EMKV-----SDIP-----QYSQ 1233
Db      1433 LEYAMKHFGELEIQATWYEKLEHWEEDALVAYDKMDTKNDOBELMGLCMRCLGALGEGWQ 1492
Qy      1234 LVANVCE--TIQEBVIALFDQTRHSLAGSATE-----DKDSMETDDCSRRHDPQRCV 1286
Db      1493 LHQCCCEKMTL-----VNDSTQAMWAMMAAAWGLGQWDSMEBYTCMIP--RDTHDGA 1544
Qy      1287 ---CVLGLH-----LAKELCEVDED-----GDSMLQVTRRL---PILPTLLTLE 1325
Db      1545 FYRAVLALHQLDLSLAQCCIDKARDLDAELTAMAGESYRAYGAMVSCNHLSELEBVIQ 1604
Qy      1326 VSL-----RMOKNLHTEATLHLLTLA-----RTOQATAVAG--- 1359
Db      1605 YKLVERREIRIQIWMERLQCCORLVEDMOKILMVRSLVSPHEDMRMLKYASLCGKSG 1664
Qy      1360 -AGITQSCILLSYUOLSTNGTACTPBARSKSLDAPBMPGY--RLSISLMEQLKTLR- 1416
Db      1665 RLALAHKTLVLILGV-----DPS---RQLDHP--LFTVHPVTAAYMKMMKRSARK 1710
Qy      1417 YNPLPEALDFVGVHQTERTQCLINAVRTVQSLACLEADHTGCF--TLQSNFMK--EMWFIH 1473
Db      1711 IDAQOHMHFVQTMQO---QAQHAIAIT-----EDQKHQELHMLMARCLYKLGEMWLN 1760
Qy      1474 L-----POLMRDIQV-----NIGYLCQACTSLHSRKMLOHLYLQNKND---GLPSA 1517
Db      1761 LOGINESTIPIKVLQYASATEHDSRWYKAMAMAMVAMFEAVLYHGYHQADDEKKLHA 1820
Qy      1518 VAQRQRPASASAPSSSKOPAADTEASFOALHT-----VOYGLKLILSTTLAL 1569
Db      1821 SGANTATATATATATATTTASTSGNSESEASTENSPTPSPLOKKTVDLSKTL--L 1878
Qy      1570 RHFTPDV 1576

```

```

Db      1879 MYTVPV 1885
RESULT 6
US-09-950-634-3
; Sequence 3, Application US/09950634
; Patent No. 6713607
GENERAL INFORMATION:
APPLICANT: Molnar-Kimber, Katherine L.
            Fallioli, Amadeo F.
            Caggiano, Thomas J.
            Nakanishi, Koji
            Chen, Yangu
TITLE OF INVENTION: EFFECTOR PROTEINS OF RAPAMYCIN
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
            Dunner, L.L.P.
            ADDRESSSEE: Finnegan, Henderson, Farabow, Garrett &
            STREET: 1300 I Street, N.W.
            CITY: Washington
            STATE: DC
            COUNTRY: USA
            ZIP: 20005-3315
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/950, 634
FILING DATE: 13-Sep-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/471,112
FILING DATE: 06-JUN-1995
APPLICATION NUMBER: US 08/384,524
FILING DATE: 13-FEB-1995
APPLICATION NUMBER: US 08/312,023
FILING DATE: 26-SEP-1995
APPLICATION NUMBER: US 08/207,975
FILING DATE: 08-MAR-1994
ATTORNEY/AGENT INFORMATION:
NAME: Slekmun, Michael T.
REGISTRATION NUMBER: 36,276
REFERENCE/DOCKET NUMBER: 01142,0058-00000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-408-4000
TELEFAX: 202-408-4400
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 2549 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULAR TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-950-634-3
Query Match 1.6%; Score 144.5; DB 4; Length 2549;
Best Local Similarity 18.6%; Pred. No. 0.0017;
Matches 336; Conservative 238; Mismatches 628; Indels 605; Gaps 87;
Qy      237 LVLTMRKEGQFGSRQTRNRLVDTMPF-----VDNIGVFSALIVEGMDIE 284
Db      217 LILTTQREPKEMQRPQWRHTFEEBAKGFDETLAKKGMNRDRI--HGALLILNELVRIIS 275
Qy      285 SLHKCALDDRRREHQPQODGLI-----CODMDCLMLTFGDIPIHAPVLIAMALLRHTINPE 340
Db      276 SMEGERL--REMEWETIQOOLVHDKYCKD---LMGFTKPRHITPTPTSF---QAVQPO 325
Qy      341 ETSVSVRKIG-----GTA----- 353
Db      326 QSNALVGLIGYSHQGLMGFTSPSPAKSTLVESRCRDLMEKFDQVCGVILKCRNSKN 385

```

```

QY 354 --IOLNFOYLTRL-----LQSLASGNDCTTS-----TACMCVYGLSPFLT 394
DB 386 SLIQMTILNWLRLRIALAFRSAPFDPOYLODPTNHALSCYKKEKERTAAFOALIGLSVAVR 445
QY 395 SLEHLTLGNQODIIDTAC-----EVLADPSLPFLFWTEPTEGSLGIID----- 438
DB 446 SEFKVYLPRVLIIIRALPRPKDFAHKQKAMQVDATVFCISMLARAMPGIQQDIKELL 505
QY 439 ---SVCGMPHLISPLDRLRALVSGSKTAKKVYFIDKMSFYNNLYGKHPDIVSHEDG 495
DB 506 EPMALVAGSPALTAVALYDLSROIPOK---KOIQGGLKMK---LSLVLMKRP----- 551
QY 436 TLMRQTPK-LIYPLGGQTNLRIPQST-VGVYMLDDRAVLVMEVSYSGSWTLF----- 546
DB 552 -LRHPGMPKGLAHQALASPEGLTTLPEASDVGSITLALRT-LGSEFEHGSILQFVHCAHD 609
QY 547 -----TCE---IEMLLHVSTADVIOHCQVKEPIDLVHKVI---STDLL 584
DB 610 FLNSENKEIRMEARCTSLRTPSIHLISGHAHVVSQTAOVVADVLSKLTVGCTDDP 669
QY 585 SIADCLLPITSRIVML-----LQRLTVISPPVDVASCNCLTV---LAARNPA----- 631
DB 670 DIRVCVLSDBERFPAHLAQAEENLOALFYALNDVFEIRBELAIC-TVGRLSMMPAFVMP 728
QY 632 -----KVTDLRHTGF-----LPRVAHPVSSLSQMSABGNMAGYGN-----L 670
DB 729 FLRKMILQILFELEHSGIGRIKEGSGARMGHLVSNAPRLRP-----VMEPLIKALIL 781
QY 671 LMNSRQGEVGVTAFLRLITTLVKQGLSGTQSGVLCVPCWFLAKEMLPSTHKKRNHSH 730
DB 782 KIKDPDPNPGVNNVATL---GEL---AQVSGL-----EM---RKM----- 816
QY 731 GVREQIGCLIELHAHILNCHETDHSHTPSLOFL---CICSLAY----- 774
DB 817 -----VDEFLIIMDLQDSSLAKQVALMTLQGLVASTGYVVEPRKXPTLLEV 867
QY 775 -----TAGQ-----TVINING-----IGVDTIDMVMAAQRPSDAEQGGQGL 813
DB 868 LINFLEKTEONQOTRREALRVGLGALDPYKHKVNIWIDQSRDASAVSISEKSDSS 927
QY 814 LIKTVKLAFSTNNVI-RLKPPSNVSP-----EOLASQHGAGNNILAVLAKIYHK-- 866
DB 928 DYSISEMVMYNGNPLDEFPYAVSNVAMLRIFRODSLSHHTVVOATTFPKSLGKCV 987
QY 867 -----HDPALPRLAIQILKRLATVAPMSVYACLGNDAAAIRDAFL--TR 908
DB 988 QFLPQVWPTFLNVRVCDGAIREFLPQGLGMLVFXSHIRPYMDEIVTLMRBEFWNNTS 1047
QY 909 LOSKIEDMRIVKVMLEFLTVAVETQ-----PGLIEFLPNEVKGSGDS-----K 953
DB 1048 IQSTI-----ILLIEQIVALGSEFKLYLPOLLPHMLRVFMHNSPGRIVSIXLAAIQ 1101
QY 954 EFSLGMSCLHAYL-----ELIDSOQODRWYCPRLHRAAIFLHALNQ--DRDSAMLVL 1007
DB 1102 LFGANLDYHLHLPLPYKLFDAPE-----APLESKRAALETVURLTESLDFDYASRII 1156
QY 1008 RTKREKMENTSPLEFGLSPSESEBSIETCALIMK----- 1045
DB 1157 H-----PIVRLTLDQSPELRSTAMDLSLVLQLGKQYIFIPMNKVLVRHIN 1205
QY 1046 -----IICLETIVVYVG-SLDDQSLKDTL----- 1067
DB 1206 HORVDVILIC---RIVKGYTLADEEBDLIYOHMLRSQGDALASGPVETGPKKLHVS 1261
QY 1068 -----KKESEIEKREPAY--WSGVYKSLAVHVAETEGS---SCTSLLE-YOML----- 1107
DB 1262 TINLQKAMGAARVSKXDWLEMLRLSLLELKXSSSPSLASCWALADAYPMARDLFMAA 1321
QY 1108 -VSAWRUL-----LIATTHADIMHLDVSVVRROFLDIVDGTALLLVPAVSN 1155
DB 1322 FVSCWSELNDDQDELIRSIETLALTSODIAEVTOITLNLLEAFCEHSHDKGPLRLDNGIY 1381

```

```

QY 1156 CLRLGSMKCTILLILLRLQMKRELGSVDEIIGPLTEILEGVLQADQCLMEKTKAVPSAFI 1215
DB 1382 LIGERRAAKCRVAYAKALHYKELEFOK-----GPPALLESISLISNNKLOQPEBA---Agv 1432
QY 1216 TVLQMK---EMKV-----SDIP-----OYSQ 1233
DB 1433 LEVAMKGFGELEIQTATVEYKLEHWEEDALVAVDKMDTNNKDDPELMLOMRMCLEALGEMQ 1492
QY 1234 LVLANVC---TLQEEVIALFPQTRHSLAGSATE-----DYDSMETDDCSSRRHDDQDV 1286
DB 1493 LHQOCCEKMTL-----VNDETOAKMARMAAAMGLQWDSMEBYCWPJ-P-RDIDGA 1544
QY 1287 ---CVLGLH-----LAKELCEVED-----GDSWLQVTRRL---PILEPTLITLLE 1325
DB 1545 FYRANVLHLDLHSLAQCCIDKARDLLDAELTMAGESISRAYGAMVSCMLSELEEVIQ 1604
QY 1326 VSL-----RMKONLHFEATLHLLTLA-----RTQCATAVAG--- 1359
DB 1605 YKLVPERRRRIROIWMERLQCGQRIVEDMQILMVRSLVVSPPHEDRTMLKVASLQCKSG 1664
QY 1360 -AGITQSCILPILSVYQLSTNGTAQTPSARSKSLDAPSWGCV-RLSMSLMEQLTKTLR- 1416
DB 1665 RLAAHKTLLVLLGV-----DPS---RQLDHP-LPTVAPQVYAYAKMKMWSARK 1710
QY 1417 YNLEPEALDFVGVOERTLQCLNAVTRVQSLACEADHTVGF-IIQLSNFMK--EWHFH 1473
DB 1711 IDAFQMHQHVQTMQO---QAQHAHAT-----EQQKHQELHKLMAKRCFLKLGKQOLN 1760
QY 1474 L-----POLMRDIQV-----NIGYLQCACTSLHSRMTLQHYLQNRKGD---GLPISA 1517
DB 1761 LOGINESTIPIKVLQYSAATEHDSRWYKAMHAWANVFEVILHYKHQONARDEKKLRHA 1820
QY 1518 VAQVORPPKASAPASAKSPQAPADTEASBQALHT-----VOYGLKILSKTLAL 1569
DB 1821 SGANITNATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1878
QY 1570 RHFTPDV 1576
DB 1879 MYTVPAV 1885

```

```

RESULT 7
US-09-538-092-1112
; Sequence 1112, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Glot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538, 092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: Curaparseqfacter Version 0.9
; SEQ ID NO 1112
; LENGTH: 2549
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc. feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number P42345
US-09-538-092-1112

```

```

Query Match 1.6%; Score 143; DB 4; Length 2549;
Best Local Similarity 17.6%; Pred. No. 0.0023;
Matches 262; Conservative 246; Mismatches 559; Indels 420; Gaps 65;
237 LVLTGKFKGQGRGQTRNRLHVDETMDPF-----VDRIQGYSAIIVGMDIE 284

```

```

Db 217 LILTTQEPKMKQKQWRHTFEEBAKGFDETLAKKEMNDDI-HGALLILNELVRIS 275
Qy 285 SLHKCALDDREELHQPADGLI-CODMCLMTFGDIPHNAPVLIAWALLRHTLNP 340
Db 276 SMEBERL-REMEBEITQOOLVHDKYCKD-LMGFTKPRHTPTPTSF-QAVOPQ 325
Qy 341 ETSSVVRKIGTATQIANVOYLTRLLOSLASGNDCTTSTACMCVYGLSLVLSLEHT 400
Db 326 QSNALVGLGYSSHQ-GLMGF 345
Qy 401 LGNOQDIIDTACEVLADSPLELPWTEPTSGLGIIIDSVCGMFPHLIS-449
Db 346 -----GTPSPAKSTIVSRCCRDLMEEKFOVOCWALYC 380
Qy 450 -----PLQLLRALVSGSTAKKAVSFLDKMSFYNELYKHPHVDISHEDGILMR 499
Db 381 RNSKNSILQMTILMLPRLAFAFBSA-FTD-TQYLOPTMNHVLSVCKEKERTAF 434
Qy 500 ROTPKLLYPLGGQTNRIPOGTQGVMLD-DRAVLYRREYSYS-SWTLFTGIE 551
Db 435 QALGLLSVAARSEKRYLPR-VLDIIRALPRKDFAKKOKAMQVATVFTC-IS 487
Qy 552 MLHVVSTADVIQHCQKVPFIIDVHKVISTDLSTADCLPITSRIYMLQRLTTVISPP 611
Db 488 MLAAQMBG-IO-QDIKELIEMLAAGLSPA-LTAVLYDSRQIOL-K 532
Qy 612 VDVIASCNCLTVLAARNPAKVTDLRHTGFLPVAAHVSSLQWISAGNAGGYNLL 671
Db 533 KDIOGGLKMLSLVLMHKP-LRHPGMPKGLAHQLASPGLTTPLEASDVG-580
Qy 672 MNSQPOGEVNTAFRLITTLVKQO-LGSTOSQVLVCWFLVKEM 718
Db 581 -----SITLARTLSFEPEGHSLTQFVRHCADHFLNSEHKIRMAAATCSLL 630
Qy 719 LPSYHKRMYNGVAREQICLILELHAII-NLCHEITLHSSHT 761
Db 631 TPSIHLSGHANVSGTAVQVADVLSLAVGITDPDPDIRCYVLSLDERFAHLAQ 690
Qy 762 PSLQPLCISLAYTEAGQTVINIGIGVDTIDMMAQPR-SDGAEQ 808
Db 691 ENLQAL-FVALNDQVEIRELALCTVGRLSNMPAFVMPFLRMILQITLEHS 744
Qy 809 GQGGLIKTVKLASVTNNVIRLKRPSNVSPLEAL-SGHAGNLLAVL 859
Db 745 GIGRIKESQAMTGLVSNAPRLIR-YMEPIKLALLIKLDPDPDENPVINNVLATI 802
Qy 860 AKYIYHKHPALPRLAQLKRLATVAPMSVYACIGNDAAIRD-AFLTRLOSKE 914
Db 803 GE-LAOVSLBMRKAVDELFIIMMLQDSSLLAKQVALW 842
Qy 915 DMRIKVMLEFLVAVETQGLIELFLNLEYKDSGSKESLSGMSCLAHLVLEIDSQ 974
Db 843 TLGQLVASTGVVEPYRYKFTLLEVLNFLTQEQGTRRAIRVGLGL-894
Qy 975 QDRVWCPRLHRAIAFLHALMODRRDSAMVLTKPKFWENTLSPLFTSPSESEP 1034
Db 895 -----DEYKHKVINGID-QSRDASAV-SLESKSSQDS 926
Qy 1035 SILETCALIMKI-ICLEIYVVGSLDQSLKDTLKKSEKRPAYMGVYKSLAVHVAE 1092
Db 927 SDYISTEMLVNMGUPLDEFYPAVSMV-ALMRFRDOSISHHTTVQAITFLFK 980
Qy 1093 TEGSSCTSLLEYQMLVSAMRMLLIATTHADIMHLLDSVVRQFLDV-1140
Db 981 SLGKCVQFLPOVM-PFLNVIRVCCDAIREPLFQOGLMVSFVKSHIR 1028
Qy 1141 --LQTKALLVPASVNCRLSGMCKTLLILIRKMKELGSVDEILPRLIEBQVQA 1198
Db 1029 PYMEIVTLMEFWWMT-SIOSTIIL-LIEIIVALLGEEFLY-LQOLIPHMLRV 1081
Qy 1199 DQOLMEKTKAVFSAFITVLQMKEMKVDIPOYSOLVNVCELTQEBEYIALFDQTRHSL-1257

```

```

Db 1082 --FMHNSPGRIVS-IKLLAALQIFGANLDDVYLHLL-PIVYKLFDAPEARLP 1130
Qy 1258 ----ALGSATEDKDSME-TDDCSRRKRDQDGVYGLHAKELCEVDEBGSWL-QV 1310
Db 1131 SRKALBTVRLTESLDFTYASRIH-PIVTLQDSPLRSTAMDTLSSLVFOL 1184
Qy 1311 TRRLPIPLTLTLEVLKMKONH-F-----EATLHLLTLARTQOAT 1355
Db 1185 GKTYQITPMWKNKLVHRRHQRVYDVLCRIYVGYTLADEEERPLVYORMLRSQGDA 1244
Qy 1356 AVAGAGITOSICLPLSVYOLSTNGTAQTPSARSKSLDAPSW-BGVYRLNSLM-EQLK 1413
Db 1245 LASGPVETG-PMKHLVSTINLQKAMGAARVSKD-DWEMLRLSLIELKQSSP 1298
Qy 1414 TLR-----YNLPEALD-FYGV-----HOERLOCLNAVRYQSLACIEBA 1453
Db 1299 SLRSCMALAQAYN-PARADLFNAAPFSCSELNEDODELIRSIELATLSODIA-EV 1353
Qy 1454 DHTVGFILQSLNFKEMHFLPOLMRDIQ-VNLG---YLQACTSLHSRKLQHYLON 1508
Db 1354 TOT-LNLAEFM-EHSDKGPLRLDDNGIVLGBRAKCRAYAKALHYKEL-EFOKG 1407
Qy 1509 KNGDLPSSAVA-QRVORPSPASAPSSSKO-PAADTEASEQOALH 1552
Db 1408 PTPALISLISINNKLOPBAAGVLEVMKHPGELFIQATWYKXH 1454

```

```

RESULT 8
PCT-US95-06722-12
Sequence 12, Application PC/TUS9506722
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: Immunosuppressant Target Proteins
NUMBER OF SEQUENCES: 25
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII (text)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/06722
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/250,795
FILING DATE: 27-MAY-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/250,795
FILING DATE: 20-DEC-1994
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 2549 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US95-06722-12

```

```

Query Match 1.6%; Score 143; DB 5; Length 2549;
Best Local Similarity 17.6%; Pred. No. 0.0023;
Matches 262; Conservative 246; Mismatches 559; Indels 420; Gaps 65;
Qy 237 LVLTLMKEQFGSGQRQNRHLVDETMPF-----VDRIQFSAIILVEGMDIE 284
Db 217 LILTTQEPKMKQKQWRHTFEEBAKGFDETLAKKEMNDDI-HGALLILNELVRIS 275
Qy 285 SLHKCALDDREELHQPADGLI-CODMCLMTFGDIPHNAPVLIAWALLRHTLNP 340
Db 276 SMEBERL-REMEBEITQOOLVHDKYCKD-LMGFTKPRHTPTPTSF-QAVOPQ 325
Qy 341 ETSSVVRKIGTATQIANVOYLTRLLOSLASGNDCTTSTACMCVYGLSLVLSLEHT 400
Db 326 QSNALVGLGYSSHQ-GLMGF 345

```

```

QY 401 LGNOODIDTACEVLADPSLPFLFWGTEPTSGIILDSVC-----GMFPHLS----- 449
Db 346 -----GTSPSPAKSTLVESRCRDMEEKFDQVCOWILKC 380
QY 450 -----PLLOLIRALVSGSKTAKKYVSLDMKSEFYNELYKRPDVIASHEDGTLMR 499
Db 381 RNSKNSLIQMTILNLPRLAARPSA-----FTD-TQYLOPTMNVHVSCKVCKERTNAF 434
QY 500 RQTPRLPLPLGGQNTLRIPOGTGVQVMD-DRAYLVREHYSIS-----SWLFTGTE 551
Db 435 QALGLSVAVSEFVYLP-----VLDIIRALPPKDFAHKROKAMQVATVFTC-IS 487
QY 552 MLHVSTADVILOHCQRYKPIIDLVHAKYSTDLSTADCLLPITSRIYMLQRLTFTVISP 611
Db 488 MLARMBG--IQ-----QDIKELLEPLMALVGLSPA-----LTAVLYLSQIQIOL---K 532
QY 612 VDVIASCVCNCLTVLAARNPAAKWTDLRHTGFLPYAHVSSLSQMSIASEGNAGGYGNL 671
Db 533 KDIDQGLKMLSLVYMHKP-----LRHPGMPKGLAHQLASPGTLTLEASDVG----- 580
QY 672 MNSEQPOGEYGVTLAFLRLITTLVYQ-----LGSTQSQGLVPCWPFVLKEM 718
Db 581 -----SITLALRTLGSFEFEHSHLTOFVRHCADHFLNSEHKEIRMAARCSRL 630
QY 719 LPSYHKWRYNSHGNREOIGCLLELIALIL-----NLCHETDLHSHT 761
Db 631 TFSHILISGHAHVVSQTAQVAVADVLSKLVAGITDPDPDIRCYVLSLDERFAHIAQA 690
QY 762 PSLQFLCISLAYTEAGQTVINIMIGVDTIDMMAAOPR-----SDGABQ 808
Db 691 ENLOAL-----PALANDQVEIRELALCTVGRLSNMFAVMPFLRMLIQLITELHS 744
QY 809 GOGOLLITKVLKASVTNNVIRLKPSPNVASPLEQAL-----SOHGAHGNLIALV 859
Db 745 GIGRIKESASMLGHVSNARLIRP--YMERIILKALIKLKDPPDPNPVINVATI 802
QY 860 AKYIYHNDPALPRLAIQLKRLATVAPMSVYACIGNDAAIRD-----AETLRLQSKIE 914
Db 803 GE-----LAQVSGLEMRKWVDELFIITIMMLQDSILARQVAM 842
QY 915 DMRIKWMLEFLTVAVENTQGLIELFLNLBYKQSDSKERSLQMSCLAHVLELISQ 974
Db 843 TLGQIVASTGVVEPYRKYPFLLEVLNLFKTEBQNOGRRRAIRVGLGLAL----- 894
QY 975 QDRYWCPELHRAAIAFLHALMODRDSAMLVLRTEKPFMENTLSPLFGTSPSEPTSEP 1034
Db 895 -----DPYKHKVIGMID-----QSRDASAV-----SISEKSSQDS 926
QY 1035 SILETCALIMKI--ICLEIYVVGKSLDOSLKTLLKFKSIEKRPAYSGYKSLAHVAE 1092
Db 927 SDVSTSEMLVVMGNLPDEFYPAVSMV-----ALMRIFRDOSLSHHHTWVQAIFIK 980
QY 1093 TEGSSCTSLLEYOMLVASAMRLIIATTHADIMHLTOSVVROLFV----- 1140
Db 981 SLGKACQVLPQVW-----PFLANVAVCGAIRHEFLFOQIGMLVSVKSHIR 1028
QY 1141 --LDGTALLIVPASVNCRLGSKCTLLILLRQMKRELGSVDEILGPTLELGLQA 1198
Db 1029 PYMDEIYTLMEFEWVMT-----SIOSTIIL-LIEQIVVALGGEFKY--LPQILPHMLRV 1081
QY 1199 DQOLMEKTKAKVPSAFTVLQMKEMKYSIDIPQYSQVLNVCETIQEIVIALFQTRHSL- 1257
Db 1082 --FVHDSRPRGIS--IKLLAAIQLQFANLDDRYHLL-----PRIVKLPAPPEARLP 1130
QY 1258 -----ALGASATEDKDSME--TDDCSRSRHRDQDGVCLGLHLAKELCEVDDGDSWL--QY 1310
Db 1131 SRKALETVRLTESLDFTYASRIH-----PIVTTDOSPELBRSTADTISLVFOL 1184
QY 1311 TRRLPIPLTLITLLEVSLEKKNLH-----FT-----EATHLALLTLARTOGAT 1355
Db 1185 GKXKQIPIPMWNKVLVHRINHQRVDVLCRIYVGYTLADEEEDPLIYQHRMRSQGDA 1244
QY 1356 AVAGAGITGOSICLPLSLVQSLTNGTAQTSASRKSLSADPSW--PGVYRLSMSLM-EQLK 1413

```

```

Db 1245 LASGPVETG---PMKHLHSTINLQKAWGAARVSD--DWLEWLRLELELKDSSP 1298
QY 1414 TLR-----YNLEPEALD-----FVGV-----HOERTLOCLNVRTVQSLACLEEA 1453
Db 1299 SLRSCWALQAQV--PWARLPLNAAFVSCWSELNBQDOLISIELALTSODIA---EV 1353
QY 1454 DRYVGFILQSLNFMKEMHFLPOLMRDIQ--VNLG---YICQACTSILSRKMLQHYLON 1508
Db 1354 TOT---LNLIAEFM-EHSDKGPPLRDNNGI VILGERAACRAVAKAHYKEL--EFQK 1407
QY 1509 KNGDGLPSAVA--QRYORPSPASAAAPSSGQ--PAADTEASEQALH 1552
Db 1408 PTAILESLSISNNKLQOPALAAQVLEFYANKHGELEIQTWYEKHL 1454

RESULT 9
US-09-914-259-11
; Sequence 11, Application us/09914259
; Patent No. 6495336
; GENERAL INFORMATION:
; APPLICANT: Makowski, Lee
; APPLICANT: Hyman, Paul
; APPLICANT: Williams, Mark
; TITLE OF INVENTION: STAGED ASSEMBLY OF NANOSTRUCTURES
; FILE REFERENCE: 8471-010-999
; CURRENT APPLICATION NUMBER: US/09/914,259
; NUMBER OF SEQ ID NOS: 180
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11
; LENGTH: 3878
; TYPE: PRN
; ORGANISM: Homo sapiens
US-09-914-259-11

Query Match 1.6%; Score 140; DB 4; Length 3878;
Best Local Similarity 18.7%; Pred. No. 0.01;
Matches 307; Conservative 232; Mismatches 583; Indels 520; Gaps 70;

QY 59 PSSAEYKANKDVASPLKELGLRISKFLGIDBOSVOLQCYQEDYRGTROSVKTVLQD 118
Db 2081 PISEHQTRVEQOLANMLKEKTDKCSLIL--LSKEQ-----LQRDIGRNEIEIKLEFR 2131
QY 119 EROSOALIIKIDYV--YEER-----TCILRCVLIHLITTYRODERHPRVYACVQDKLE 172
Db 2132 VRELEQALVSADTFQKVEDRKHFGAVEAKPELSLEVOQLAER-----DAIDRKEK 2183
QY 173 LVSKYRQ--QF--EELYKTEAPTWEHGNLMTEROVSRWFVQCLREQSMLEIIFLYAYF 229
Db 2184 ITVLEQQLBQFRELBNKNEBVQQLHMQLEIQKSTRTLOLEBQENKLFK----- 2234
QY 230 EWAAPSLVLVTMFEQGGSGRQNRHLVDETMDFVDRIGYSALIVGMDIESLH-- 287
Db 2235 ---DMERKLGALIKESDAMSTQ--DOHVL-----FGKFAQIIOKEVEIIDLNQ 2279
QY 288 -----KCALDDR--REIHOFAQDGLICQDMDCLMTFGDIPHAPVILAMALLRHTL 337
Db 2280 VTKLQOQLKITTDNVIKNEKLIRD--LEIQICLM----- 2314
QY 338 NPEBTSSVVRKIGGTALQLVNFQVYLRILQSLASGANDCTSTACWCYGLLSFVLTSL 397
Db 2315 --SDQCVKRNNEEIEQLN--EVIEKLQOELANIGQK-----TSWN 2352
QY 398 LHTLGNQODIITDPAQVYLDAPSLPLFWTEPTSGIILDSVCGMFPHLSPILQOLRA 457
Db 2353 AHSLSHADSLKHQDLVAVAEKL--ALBQCVETANEBMTFMKVNALKETNKMQULTQELFS 2411
QY 458 LVSGKSTAKVYSF-----LDKMSFYNELYKRPKPH-DVISHBDGLMRQRPGLALVP 508
Db 2412 LKRRESVYKISQISIPENSUNVAIDLHS-----KQKPEFLVLTEDA-----LKS 2455
QY 509 LGGQTNLRI--POGTGVQVMDDRAYLVREHYSISWTLFTCEIEMLLHVSTADVILOHC 566

```

```

Db      2456 LENQYFYSFEENGKGINLETR-----LQLEST----- 2486
Qy      567 QAVKPIIDLVKAVITDLSIADCLPI-----TSRIYMLQRLTTVISPPVDVIA 616
Db      2487 -----VSAKDLELTQCYQIKDMQEOGFETEMLOKXIVNLQKIVEE--KVAA 2532
Qy      617 SCVNCLTVLAARNPAKWTDLNHTGFLPVPVAPVSSLSQMSAEGNAGVGNLMSNSQ 676
Db      2533 ALVSGIQLEAVQEVYKFCOD-----NOTTISEPRT-----NIONNL 2571
Qy      677 POGEGVTIAFLRLITTVKQ-----LSTQSGLVPCVFMFLKEMLPYHKMRY 727
Db      2572 REDELGDISALTALNISELSEQVEMHTSLLEKQVEIAEKXVLEKEKKLELQKLEGG 2631
Qy      728 NSHGVE---QIGCILELILHAILMCHETDLSHSHTSPLOFLCISLA-----YEA 777
Db      2632 NEKKQREKEKRSPODVEVLKTTTELPHSNE-ESGFEMLEALRAESVATKALASYEK 2690
Qy      778 GOTVININGIGVDTIDMWAAQPRSDGALGQGGOLITVVALSPVNNVIRLKPSPV 837
Db      2691 AEKL-----QEBLVNE-----TN----- 2704
Qy      838 VSPLEALSQ---HGAGNNLAVLAK-----YIYHKDPALPRLAIQLKELATVAPWS 889
Db      2705 MTSLOKDLQVARDHLAEAKKSLILEKDETEVQSKKACWEPPLPYKLSKSIASQTOGT 2764
Qy      880 VYACIGNDAAI---RDAFLTRLOSKIEBMRKIMLEFLVAVETOPCLIFLNLKY 946
Db      2765 LKISSNOTPOLVKNAGIQIMIOGECSESEVTEIISOPTKEIKEMQELHAAEILDMESR 2824
Qy      947 DSGDS--SKERSLGM-----WSCLHAYLELIDQOQORW--CPPLHRA-----AI 989
Db      2825 HISSETLKEHYAVVOLKEECGTAKAYIQCLRSSEVGFYNNCPSTLCSGSMGCI 2884
Qy      990 AFLHALMOD---RRDSAMVLRTKPKFWMNLTSPLFG-----TLSPPS----- 1029
Db      2885 YLTHSGQDIASEGGESESATDSFPKKIKGLLAVNHMGQVLSLSTEPSDDEHSI 2944
Qy      1030 -ETSEPSLLETICALIMKIICTEIVYVKGSLDOSLKDITLKRSITEKRPAYMSGYKSLAV 1088
Db      2945 QOVSEPMLEERKAYINTI-----SSLKDLITKMOQORE----- 2977
Qy      1089 HVAEBSGSSCTSLSEYOMLVSAWMLIITATHADIMHLDTSVVARQLFLVDGTKALL 1148
Db      2978 --AEVYDS--QSHESFSDMRGELLAL-----QOVFL-----ERSVL 3012
Qy      1149 LVPASVNCIRLGSMTCTLLILLRQMKRELGSVDEILGPLEILEGV--LOADQOLMEKTK 1207
Db      3013 LAAPTELTLALGTDAVGLNLEORIG-----QGEVQAMECLOKAD 3057
Qy      1208 AKVESAFITVL--QKEMKV-----SDIPQSVLVNVCETLOBEVIALFDQTRHSIA 1258
Db      3058 RRSLSSEIOALHAKMGRKITLKEQSEKPEQELLEVI--QOKOSOMLEM--QVELSSM 3114
Qy      1259 LGSATEDDSMETDCSRSRHRDQDGVCVGLMLAKELCEVDDEGDSMLQVTRRLPIIP 1318
Db      3115 KDRATELOEOLSEK-----KVVAELKSELAQ----- 3141
Qy      1319 TLLTTLVLSLPMK--ONLHTEA-----TLHLL--TLARTQOGATAVAGITQ 1364
Db      3142 ---TLELEJTLKQHKHKLKELEAFRLVEXKTDVHLLNDLTLAEQKSRRLQALRE 3198
Qy      1365 SICPLVLYOUSTNGTAOTPSARKSLDAPSPGVYRLMSLMQOLKTLRYNLPAL 1424
Db      3199 KAKL-----GRSEERDKBELE-----DLKFSLESQOKPMLQNLNL--- 3233
Qy      1425 DFVGNHQBRTLOCLN-AVRTVQSLACLEADHTVGFIQLSNFMKEMHNLPLQMLRDIQV 1483
Db      3234 -----LEQOKLMSQOKIESQMLYDA-----QLSEQGR--NIELQVLLSESEV 3278
Qy      1484 NLGYLQACSTSLHSRKMLOHLYLQNKNGDGLPSAVALORVORPPSASAASSKOPADT 1543

```

```

Db      3279 RI-----REMSSTLDERELHQAQSSDGTG-----QSRPP-----LPSEDL 3315
Qy      1544 EASEQALHTVOYGLKLKST 1565
Db      3316 LKEIQKQKEKRSRIVELLNET 3337

RESULT 10
US-09-270-767-44725
; Sequence 44725, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7325-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; CURRENT FILING DATE: 1999-03-17
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 44725
; LENGTH: 919
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-44725

Query Match      1.5%; Score 136.5; DB 4; Length 919;
Beet Local Similarity 18.9%; Pred. No. 0.0015;
Matches 173; Conservative 134; Mismatches 322; Indels 287; Gaps 44;

Qy      944 EVKQSDSDSKERSLGMSCLHVALLELIDQOQDRYWCPLLRRAAIFLH-----ALMQ 997
Db      30 EKKQDKDEQ-----PELSBDQ-----LQEBLLVGHLOEPDAKLQ 70
Qy      998 DRDSAMVLRTKPKFWMNLTSPL-----FGLTSP-----PSETSEPSILETCALIMKI 1046
Db      71 PTLSEMAKLIRASTYMTSVKPLFKMRPHYETMTKYKHMENEQAR--QLCADIIISV 126
Qy      1047 ICLTEIVYVKGSLDOSLKDITLK--KFSIEK--RFAYWS--GYVSL-----AVHVAEBSGSSC 1098
Db      127 LSWTV-----GSGKDCIAYRFLCDKQKIGMGBHYVHLLSGEISAHYHDTTGDR 177
Qy      1099 TSLSEYOMLVSAWML-----LTIATTHADIMHLDTSVVARQLFLVDGTKALL 1148
Db      178 VQLELVYQIIPYNNHNAADACDLIT-----EIDHLLHLSFVDESAYPR----- 225
Qy      1149 LVPASVNCIRLGSMTCTLLILLRQMKRELGSVDE-----ILGPLEILE 1193
Db      226 -----CLYLOS--CYPVPDPDNTIILETALQLSRKFNQYQAMRLAMDMDKIGE 276
Qy      1194 GVLQADQOLMEKTKAKVSAFITVLQKEMKVSDFQYSQLVLANVCETLOBEVIALFDQTRHSIA 1253
Db      277 IFKEKEBPAMQQLAFMLARQOICLELDEL-----VPDQDILM-----EIMSNANLN 323
Qy      1254 RSHLSLGSATEDDSMETDCSRSRHRDQDGVCVGLMLAKELCE----- 1299
Db      324 KHFLNLARELDIMEKTEBEDIKSHLDSRSRFAIYQVDSAKQNLAAFPVNGFVNAFGCV 383
Qy      1300 ---VDEGDSMLQVTRRLPIIPT-----LITTLVLSLPM--KONLHTEATL----- 1341
Db      384 DKLSLEDGNKMLYKXKKEGMLSATASLGLILIMVDVGLTMIDKLYSTDNDIKSGALIA 443
Qy      1342 -----HLLTLARTQOGA-----TAVVAGITQSIGCLPL--SYQOL 1376
Db      444 CGIYVNGIRNEVDPAHALISDYIDNONSOMRGVAILGLIAGVAGNSRSIVDTLKTVPSE 503
Qy      1377 STNGTAQTPSARSKSLDAPSPGVYRLMSLM-----EQDLTKTL----- 1415
Db      504 GSNMNS---SASVEL-----GITALSIGLIVSGVSCNSSEITELILQITMGTAKDLMDT 554
Qy      1416 --RYNFLPEALDFVGNHQBRTLOCLNAVRYVQSLACLEADHTVGFI-----QLSNF 1466
Db      555 YTRFLFLGLIGLYIG-RQKST---EAV--MMTLEVLLEBPYKSNATVWDICAVAGTGNV 607

```



```

Db      1948  TNEP---HLTLEFEESIQGF-----QRTTELEKX-----LCLF-----1978
Qy      1455  HTVGFILQISNFMKWHHLPOLMARDIOVNIQYLCOACTSLHSHKMLQHTLQNKNGGL 1514
Db      1979  -----YMTW-----LKNLKFCKSNDSDSKLKYVQSLDK-----L 2009
Qy      1515  PSAAVQRVPRPSAASAPSSSKOPADTEASEQALH-TVOYGL-----LKLSTLAL 1569
Db      2010  INLTIDQKEMTPSVQAKIWGSIQILPELIDWVLDNPLHSTYGGSGPQVEIMADTAYAL 2069
Qy      1570  -----RHFTPDVCOIL-----LDQSL---DLAEY-NFLPALSETTPTPDS 1606
Db      2070  ASANQVLSKVIYTRICVMDSCNTPVOYLEQHMMMDIALGRVYLMLSFNCL---D 2126
Qy      1607  VAPSEGTLLATV-----NVALNMLGELDKKEPELTQAVGLSTQAEGRRT 1650
Db      2127  VATSPVYLFHTITFLVCSGSLMRASTHGLVINIHSICTCTNP-----SFSEARV 2179
Qy      1651  LK-SLMTMENCFLYLI-----SQAMRYLDRPAVHPRDK---QRMKQELSSLSLTL- 1698
Db      2180  LRLSLDEFSLPK-FYLLRFGISKVSAVAATAFRSGCRHPTDKMLGNERVOTPLPADRELS 2238
Qy      1699  LSSL 1702
Db      2239  LPSL 2242

```

```

RESULT 12
US-09-510-791-1
; Sequence 1, Application US/09510791
; Patent No. 6365126
; GENERAL INFORMATION:
; APPLICANT: Zhong, Yi
; APPLICANT: Guo, Hui-Fu
; APPLICANT: Tong, JiaYuan
; TITLE OF INVENTION: Improvement of Learning and Short Term
; FILE REFERENCE: 1314, 1047-002
; CURRENT APPLICATION NUMBER: US/09/510,791
; PRIOR FILING DATE: 2000-02-23
; PRIOR APPLICATION NUMBER: US 09/046,745
; PRIOR FILING DATE: 1998-03-24
; PRIOR APPLICATION NUMBER: US 60/041,469
; PRIOR FILING DATE: 1997-03-24
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 2802
; TYPE: PRT
; ORGANISM: Drosophila
US-09-510-791-1

```

```

Query Match      1.5%; Score 136; DB 3; Length 2802;
Best Local Similarity 18.8%; Pred. No. 0.014;
Matches 339; Conservative 244; Mismatches 635; Indels 586; Gaps 93;

```

```

Qy      169  LEKELVSKYRQOFEELYTEAFTW-ETNGNINTERQVSRWFOCLREOSMLLEITFLYYA 227
Db      755  LOKRIMTLIR-KIEHCVAGVQPAWETFRNWEVSSKVLQTYPKCGEDQO----- 803
Qy      228  YFEAPASLLVLTKFKKQGRSR---QTRHLVDENMDPRVNDIGVFSMLILVEGMDI 283
Db      804  -----AEVF-HRGMGKRASHQSSHEDEQINEMANTWFLALGVS----- 845
Qy      284  ESLHKALDRRELHQFADQGLICQDMCLMTFGDIPHHAPVLLAMMLRLHTLPEETS 343
Db      846  -CLHGRS-SRQMLQOSQNN-----ASLG-----SLAQNSLYSSSTS 881
Qy      344  SVRKIGGTAIQNL-----VFOYTLRLQLSLASG-----G 373
Db      882  SGHGLSHPSVSLSTLPPAPPODVSYCPVQGLRLRLVCSNEKIGLQIKVKEILVG 941

```

```

Qy      374  NDCTTSTACMCVYGLL-----SFVLTSLELHTLGNQODIIDTA 411
Db      942  EEMSTQ-----LYPLFPOVAIYEKFPDQOQVAVNVNTDINTQIERTIYIMKSIDLPK 996
Qy      412  CEVLDPSELPELFMGTEPTSGIILDSVCGMF-----PDL-----LSPLQLL 455
Db      997  AN--KDPANNDO-----PSPSEHLGV--TSIEGMGLIYRVHRLDMVTVAIRIKTYLQCLV 1048
Qy      456  RALVSGKSTAKKIVSFLDKMSFYNELYKHKRHVDVSHEDGTLNRQTKLYPLQGO--- 512
Db      1049  EVMKMRD-----LAFRQEMKFRNKLVEYL-----TDWVGSHQIAPRESSADAA 1094
Qy      513  --TNLRIPQGVQVQMLDRAYLVWREYSSSWTLFTCEIEMLHVSTADVIQHCQRYK 570
Db      1095  ILTNTSLIFRDLDAQCAVALLIR-----GLPLQPESSRGDLMD---AK 1137
Qy      571  PIIDVHKVISTDLSIADCLPITSRIYMLQRLTVVISPVDVASCNCL---TVLAA 627
Db      1138  SALFLKYFTLPMNL-LINDCISSEAKEM---NNTPLPRPPMAAGRLTLRNATILAM 1193
Qy      628  RN--PAKYTDLRHT--GFLPVP--AHVYSLSQMS-----AEGMAGGYNL- 670
Db      1194  SNLGANIDSGIMHSIDGYNPDLOTRAAFMEVLTOIQOGTEPDTLAEVLADEFQLV 1253
Qy      671  -LANSQOPQGEYGVTAFLRLITTLVKGLSTQSGGLVPCVMFYLKEML-PSYHKMRYN 728
Db      1254  QVTWISPKGELPIAMALANVYTSQMDLAK-----LVYTLFPAKILSLBLMMMYR 1307
Qy      729  SHGVREQICLILBELIHAILMLCHETDLSSHSTSLQFLCISLAYTEAGOTVINIMIG 788
Db      1308  EWEVSD-----CMQT-LFRGNSLGSKIMARCFKLY-----G 1337
Qy      789  VDTIDVMAAPRSDGAEQOGQLITKYKLASVYTNVNRILKPSNVVSPLEQALSQH 848
Db      1338  ASYLQMLLEPLIR-----PLDDEEBTECEV--DPAALPTEDI---ED----- 1376
Qy      849  GAHGNMLIAVLAKY---LYHGHDPALPRLALO-----LAKRLATVAPMSVYACIADA 898
Db      1377  -HRRNLIALQKYPDAIINSDDRPPQJBSMCHGLYVLSKRPNL-----LQNNI 1426
Qy      899  AAIRDAFLTRL-----OSKIEDMKIKMILEFLTVAVETQGLIEL 939
Db      1427  GAVGVVILRFLTNPIAVSPQELGIYDKOVHSSAKRGIMLMSKILQNIANHVFSKEQML 1486
Qy      940  FLNLEVQKDSGSKSEFSLGMS-CLHAYLELIDSGQDQRYWCPRLLHQAALFLH-AW- 996
Db      1487  CFNDFLRHFAGRFFLIQIASDC-----ETVDQTSHEM---SPISDANYLALHRLMT 1537
Qy      997  -----QDRPSAMVLRTKPKFWENLSPLEGLTSPSETESEPSILETCALIMKLI 1047
Db      1538  HOEKIGDYLSSSRDHKAVGRPRFDM-----ATLLAYLGPEBKFEVDSHMFSS----- 1586
Qy      1048  CLEIYVVVKGSLDQSLKD-----TLKKFSI-----EKRPAYWSGYVKS 1085
Db      1587  -----YAWSSIDMSSTWFEIIVWQHMEKEBEFTLKSNNIIFQAGTSKSGYPVFIYA 1641
Qy      1086  LAVHVAETEGSSCTSLLEYQMLV-----SARMLLIATTHADIMLTLSVVRQQLFL 1138
Db      1642  RRYKIGETNG--DLIYHVLTLKPCHPFEVVDFTHTCSNDRFRTEFLQMFVYL 1697
Qy      1139  DVL--DGTKALLVPAVNSGRLGSMKCTLLILLRQKRELGVS-DEILGPL----- 1188
Db      1698  PTVAVENNAVYI---YNC-----NSWVREYTFKHRIILAPLKNRKL 1738
Qy      1189  -----TEILEGVLOADOQ-----LMEKTKAVFSAFI-----TVLQMKEM 1223
Db      1739  FLESNPKLTDFPILABQQLPGATILSLDEDLKVFVSALTLSHKQTKVAIKVPTALQITSA 1798
Qy      1224  KVSIDPQYSQVLAVNCETLQSEVALPQTQTHSLALSGATDKDQSMETDDCSRSRHRQR 1283
Db      1799  EKTIVLAHSVILINDVYASIEBVCVLDNDNOFTLSTNES--GOLSFHNDC----- 1848
Qy      1284  DGVCVLGIHLAK--ELCEVDEDDGSMQLQVTRRLPLP-----TLTLTLEVS-LRMKQNLH 1335

```

```

Db      1849 DNIYQAIHHRMELSDPD-----STYVQKIRPKVPGTILMALINIGSCDPNLR 1901
Qy      1336 FTEATLHLLLTARTOGATVAVAGI-TQSTICPLISVYQUSSTNGTAQTPSARKSDA 1394
Db      1902 --TAAYNLLCALATAPD--LKIEGQLLETQGLCP-----SNNTIFIKVSEKL--A 1947
Qy      1395 PMPGVYRLSMGLMEQLKTLRYNPLPBLADPVVGHQERTLOCLNAVRTVOSLACLEAD 1454
Db      1948 TMEP---HTLLEFLERSTIGF-----QRTTLEKQ-----LCLF--- 1978
Qy      1455 HTVGFILQLSNFMKWHFHLPOLMRDIOVNLGYLCOACTSLHSRKMLOHYLONKNGDGL 1514
Db      1979 -----YMTPW-----LKNLVKFCNSNDSSKKLKVSQLDK-----L 2009
Qy      1515 PSAVAGRVQRPSSAASAPSSSKQPADTSEQQALH-TVYGL-----LKLKLTAL 1569
Db      2010 INLTIDQKEMVPSVQAKTWGSIQIPELIDWLDNLFHKSITYGSPQVEIMADTAVAL 2069
Qy      1570 -----RHFTPDVCOIL-----LDQSL---DLAEY-NFLPALSTPTPTPDS 1606
Db      2070 ASANYQVSKVITITICRMWDSCTNPQVYLEQHMMMDIALGRYLMLSFNCL---D 2126
Qy      1607 VAPSGTLLATV-----NVALNMLGELDKKEPLTQAVGLSTQAGTPT 1650
Db      2127 VATSVVPEFHTITPLVCSGSLSMRASTHGLVINIHSICTCTNP-----SFSEAGRV 2179
Qy      1651 LK-SLMTFMCNCFYLLI-----SOAMRYLDPVAVPRDK---QRMKQELSSLSLT- 1698
Db      2180 LRLSLDEFSLPK-FYLLFGISKVKSAVATARRSSCRHFTDKLGNERTVQPLPADRELS 2238
Qy      1699 LSSL 1702
Db      2239 LPSL 2242

```

## RESULT 13

```

US-09-270-767-33392
; Sequence 33392, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 33392
; LENGTH: 147
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-33392

```

```

Query Match      1.5%; Score 131.5; DB 4; Length 147;
Best Local Similarity 29.7%; Pred. No. 0.00015;
Matches 41; Conservative 25; Mismatches 61; Indels 11; Gaps 5;
Qy      898 AAARDAFLRLQSKIEDMRKIMLEFLVAVETQGLIELPLN---LEVKDGSQSK 953
Db      1 ADQIRLFEMQKLPDELESDSIKIATLELVACIAKQPGVTEAFKXVAVLADKRSRFFSK 60
Qy      954 EF--SLGWSGLHAVLELIDSOODRYWCPLHRAAIAFLHLMQDRDSAMLV--LRT 1009
Db      61 DCVENIGS-STVYMRPFLDALQVDPLTIOQALPAKIMTIFHSWM--KHNLOMLVDDLVK 117
Qy      1010 KPKFMENTSPPLFGTSLP 1027
Db      118 DKQFWKLCSPFLFSELQ 135

```

RESULT 14  
US-09-270-767-48609

```

; Sequence 48609, Application US/09270767
; Patent No. 6703491
; GENERAL INFORMATION:
; APPLICANT: Homburger et al.
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster
; FILE REFERENCE: File Reference: 7326-094
; CURRENT APPLICATION NUMBER: US/09/270,767
; NUMBER OF SEQ ID NOS: 62517
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 48609
; LENGTH: 147
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
US-09-270-767-48609

```

```

Query Match      1.5%; Score 131.5; DB 4; Length 147;
Best Local Similarity 29.7%; Pred. No. 0.00015;
Matches 41; Conservative 25; Mismatches 61; Indels 11; Gaps 5;
Qy      898 AAARDAFLRLQSKIEDMRKIMLEFLVAVETQGLIELPLN---LEVKDGSQSK 953
Db      1 ADQIRLFEMQKLPDELESDSIKIATLELVACIAKQPGVTEAFKXVAVLADKRSRFFSK 60
Qy      954 EF--SLGWSGLHAVLELIDSOODRYWCPLHRAAIAFLHLMQDRDSAMLV--LRT 1009
Db      61 DCVENIGS-STVYMRPFLDALQVDPLTIOQALPAKIMTIFHSWM--KHNLOMLVDDLVK 117
Qy      1010 KPKFMENTSPPLFGTSLP 1027
Db      118 DKQFWKLCSPFLFSELQ 135

```

## RESULT 15

```

US-09-540-236-3580
; Sequence 3580, Application US/09540236
; Patent No. 6673910
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO MORAXELLA CATAR
; FILE REFERENCE: 2709.2005-001
; CURRENT APPLICATION NUMBER: US/09/540,236
; NUMBER OF SEQ ID NOS: 2000-04-04
; SEQ ID NO 3580
; LENGTH: 1383
; TYPE: PRT
; ORGANISM: M.catarhalis
US-09-540-236-3580

```

```

Query Match      1.4%; Score 130.5; DB 4; Length 1383;
Best Local Similarity 18.9%; Pred. No. 0.013;
Matches 278; Conservative 200; Mismatches 513; Indels 483; Gaps 68;
Qy      81 RISKFL-GLDEGSVOLLQCYQEDYGRGTRSVKTVLQDERQS--CALILKADYYEER 137
Db      119 RIFGFIKTLSEQDIKIXI---MQDDHHPYDFIMTIYDTQNGSIPRHRLWRICE--EVS 171
Qy      138 TCIIRCV-----LHLLTY-----FQDERHRYREYADCVKKEKELVSKY---- 177
Db      172 SVFVRVYLTRPEWHLAMTHDPLPESIDEMITDKORFANDYSDS-DPTKMLIQYHDL 230
Qy      178 -----ROOFELXY---KTEAPTW-----ETHGNLMTERVQVSRW 207
Db      231 KILRYLMRQLFSVAVYRFTLEVRFAVALDGLRGSSVAEAYKSLPNHLYLPTIOQIPL 290
Qy      208 FVQCLREQSMLEIITLYYAYFEVAPSDV-----LVTXKF 243
Db      291 ELQFLRLSVHNLVNLHFNPSKMFWDIVDKRWLATORIINPNVYLLKDYGHVLSRLG 350
Qy      244 KEQFGSRQTNHNLVDETMDPFVDRIGVFSALILVEGMDIESLHKCALDDR--ELHQRA 301

```



```

Db 351 KE-----SRFETAMLADMSGDY-----YYEK-----EGV-----NRRFETVTRTP 386
Qy 302 QDGLICQMDCLMTFGDI PHHAPVLLAMALLRHLNBEETSSVVRKIGTALQULNVQY 361
Db 387 KDWQWQDDDFV-----SPAQESKSLTN-----OLKSDIIMLBDEGATKQWQOE 431
Qy 362 LTRLLQSLASGNDCTTSTACMCVYGL-----LSFVLTSLHETLGNODDIITACEVLAD 417
Db 432 LVSAFKNSAFWEQAKTKTSQMLPLGLQKNDOSLRLPSLSIHACHSLKQLEEVAMITAR 491
Qy 418 PSLPELFWTEPTSGIILDSVCGMPHLL-----SPLQL-----LR 456
Db 492 -----YLNBNPS-----DQSCRLKSDIVVYLPEAQAEDLRLIENDVGMGLKLP 538
Qy 457 ALVSGKSTAKKYVSLDKMSPTNEIYKXKPHDVISHEGOTLKRQTPPLPLPGQOTMR 516
Db 539 AKITG-TASREVDSILMAAISGFYRLGAK-NARFYRDVYEWMTPLVYSLG-----LD 591
Qy 517 IPQGVGVMLD-----DRAVLR-----WEYSYSWTLPTCEIEMLLHVSTADV 562
Db 592 PESVNRACQLDAGFRGFDAAHLAMTLNADDIRS-----PSYALDRIV-----AGF 642
Qy 563 IQHCQRVPIIDLVHKVISTDLSTADCLLPITSRIYMLQRLTVVISPVDV----- 614
Db 643 LTPCDSFSP-SLMLHPFKMRDESFQESLP-----LEGVTLADQIVLEALCLIHAGL 693
Qy 615 -----IASCVNCLTVLAARNPAKWTDLRHNGFRLPAVAPVSSLSQMSAEGNANG 665
Db 694 REFDEYORLAPVLDMLHIEHRVNRIFYTKYRETSQMHVIFKWNGLIASIRAN-----H 749
Qy 666 GYGNILMNSQEPQGEYGTIAF-LRLITLVKGQSGTOSQGLVPC-----VMFYLK 716
Db 750 HYHHTKNSGTQNTQIETLEFVLESMASSQVSAFRACTIFPAFGALRVNPFGLT 809
Qy 717 EMLPSYHKRYSNGVREQIGCLILELHAILMLCHETDLSSHSTPS--QFLC---ICS 771
Db 810 LMLD-----NMLSAFPRQDQARLDLMAKGLKRRGRFLEDDDGAFDLAILGSRDCA 863
Qy 772 LAYTEAGCTVINIMIGVDTIDMWMAAOPRSDGAGQOGOLLITVYKLAASVTNNVIRL 831
Db 864 IFYTNV-----ADDTGOLLRPASVSB-----LIEFFK----- 890
Qy 832 KPSNVVSLPEOALSQGHAGNNLAVLAKYIYHGNDRALPRALQOLKRLATVAPMSVY 891
Db 891 ---SNV-----DWAYDLOADHTNQ-----ADQIAKALATTIAPKIVQ 923
Qy 892 ACLGNDAAAIRDAPLRLQSKIEDMRI-KVMLEFLTVAVETOPGLIEFLINLEVKQSD 950
Db 924 QLLIYEHNA--NAPDASVFETKEWVCDEDDLADPLQKI----- 961
Qy 951 GSKERSLGMSCLAHVALELISQODRWCPPLHRAALAFHLMQDRDSAMLVLRK 1010
Db 962 -----LHA-----KOSQTYLPR--SPMNVKLRHLLDOFPNOSPISL-PD 999
Qy 1011 PKFENLTPFGLTSPPS-ETSEPSILETCALIKWICLSEIYVVKSLDQSLKDTLK- 1068
Db 1000 ATTYQMLABALQNLQSGQLEEVLTITGLOLPRMSPLNOIYHKLKPARAYLHKLPL 1059
Qy 1069 KFSIEKRPAYMSGVYKSLAHVALEGGSCSTLEYOMLVSAWRMLLIATTHA----- 1122
Db 1060 LYDABEDGFL-----EERPLSDLOGGYOL-----KETLIDMAAQGVFDGMP 1099
Qy 1123 -----DIMHLTDS-----VPRQLPLVDLDTGKALLVPAVNVCLRLGSMKCT 1165
Db 1100 MDEPDYHKLITDIYNNDDKASEFAHIVR-IYYD-----NLLPAQAN----- 1139
Qy 1166 LLLILRLQWKRELVSEILGPLTEILGVALQADQOLMEKTKAKYFSAFI--TVLOMKEM 1223
Db 1140 -----RLTSLDNLVGLAKLTDTPAONLTDLOVKKQOLOYSMISPTKQYIAY 1188
Qy 1224 KVSDBP-----QYSQVLVANCET-----LOEVIYALFDQTRHSLAGS 1261
Db 1189 ELNHLPLDIYVAKLPLISCDQMLRVLNPTVRAGHFLAFYLHLHAWQSRQTDQDIDRQNLGS 1248

```

```

Qy 1262 -----ATEDXDSMETDDCS-----RSNRD-----QRDVCVGLGHLAKEICEVEDGDS 1306
Db 1249 SLWQFGQNSNDLKDPHLNDVSLKLKPIHKETAKRQLOENFWMG-LAKQ----- 1297
Qy 1307 WLQVTRRLPILPTLLTLE-VSLRMKONIAFTEA 1339
Db 1298 -----MPVLPLMTSLNIVALTNNLEBEGES 1324

RESULT 16
US-09-198-452A-113
; Sequence 113, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:
; APPLICANT: Griffiths, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, preve
; TITLE OF INVENTION: and treatment of infection
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/09/198,452A
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 113
; LENGTH: 1826
; TYPE: PRN
; ORGANISM: Chlamydia pneumoniae
; FEATURE:
; NAME/KEY: SITE
; LOCATION: 1...1826
; OTHER INFORMATION: Xaa=unknown or other
US-09-198-452A-113

Query Match 1.4%; Score 129.5; DB 4; Length 1826;
Best Local Similarity 18.8%; Pred. No. 0.028;
Matches 360; Conservative 233; Mismatches 627; Indels 695; Gaps 93;

Qy 223 FLVYAFPMASDILLV--LTQMFKEQGRSGFQTRHNLVDETMDPVDVRIGVFSALLIVEG 280
Db 92 FSHYSHAVTSGSTTELFSHALLFTLEG---QADPKTEVLDDLYSKERKVLSTWELSEG 147
Qy 281 MDISLHKCALDDBRELHQAQDGLICQMDCLMTFGDI PHHAPVLLAMALLRHTLNP 340
Db 148 VQISLAPLARKDLAITHVAAQO-----FTKVCNGIHIHYFSLTSGI--PE 194
Qy 341 ETSSVVRKIGTPI--QLNVFOYLRLQLSLASGNDCTTSTACMCVYGLSPVLTSL 398
Db 195 DCS--VDIVIDTLIKSENNIARLKVSLPTALEFGHCS-----VLSDEL 238
Qy 399 HTLGNQDIIPTACEVLADPSLPELFWCTEPTSGIILDSVC-----GMFPHLSPL 452
Db 239 MTFSTKQDID---VYTPPLTQOLF--SPHA-----LESRCSLQSGSGIFISIDNPL 286
Qy 453 QLRBALVSGKSTAKKYVSLDKMS--FYNELYKXKPHDVISHEDGTLRROTPKL--LYP 508
Db 287 -----IDENLSIKENCCSPANCSSYLXHTIYQALA-DALFNLLETPKXDSLPEIQNIF- 339
Qy 509 LGQOTNLRIPQGVGVN----- 526
Db 340 LRGNKNTLVPRLPDQTLGKKNLTYYKVRGVLANDIGDKVRYTTPRSRYLSKMSAHSCSL 399
Qy 527 -----LDDRALVWNE-----YSYSWTLPTCEIEMLLHVSTADVQHC-QRYKP 571
Db 400 CKGTGLGQYASVAVTEGKTFTEFOQMSLNHVFSSKVS--PSLSIOELIQLKQRLSP 457
Qy 572 IIDLHVHKVISTDLSDAC-----LPTSRIVMLQRLTVVISP-----VD 613
Db 458 LIDGLGLVLTNRRALATLSGQERTIAKLHGLGELPRTIYIILDEPSLGLHROPQTEKIG 517
Qy 614 VIASCNCLTVLAARNPAKWTDLRHNGFRLPAVAPVSSLSQMSAEGNANGYG----- 668
Db 518 VI-----KKLRDQGNFVILVEHEERMISLADRIIIDIGPAGIGGGEVLF 561

```

QY 669 -----NLNNSPOGEY-----GVTIAFLRL 690  
 Db 562 NKRPEDFLNNSSSLTAKYLRQELTPIPSREAPISWLLTATITNKNLSIRPLAFL 621  
 QY 691 ITTLVKGQSGTOS-----OGLVPCWFLKEMLP--YAKWRVNSHGVEQIGCLILELI 744  
 Db 622 IG--VTGVSGSGKSLINNTLVPAIBSFLOENPKMLHFEW-----GC-IGRLI 667  
 QY 745 HAILNCHETDLHSHSTPSLOFLCICSLAYTAGQVINIMIGVDTTDMMAAPRS-- 802  
 Db 668 HI-----TRDLPGSQRSI-----PLTYIKA-----PDDIRELPASQPRSLR 704  
 QY 803 -----DGA--EGOGGOLLKTVKLAFTVNNVIRLKPNSVSPLE-----Q 843  
 Db 705 OGLTKAHSPFNQPGACICOGGLGIMTI-----SDDTPIPCSECO 745  
 QY 844 ALSQGA-----HGNMLIAVLAKYIY-----HKHDPALPRIAIQLKRLATV 885  
 Db 746 GKRYHSEVLEILYEGKNIDILDMTAYEAKFPFISHPKIHEKIALCGLRID----- 797  
 QY 886 APMVYACIGNDAAIRDAFLRLSKIE---DMRKMILFELTAVETQ--GLIEL 939  
 Db 798 -----YLPGRPLSTLSGEILQRLKLAHELLFASPKQTLVYLDPTGLHTHDIOALIEV 852  
 QY 940 FLNLEVKDGSDESKESFSGMWSCLHAYVELIDSOQODRYWCPPLHRAAIA-FLHALMOD 998  
 Db 853 LLSTLY-----LG-----HTVLYIEHN-----MHVVKCDVIELGPE 885  
 QY 999 RRD-SAMVLVTRKPKFMENTLSPFLGTSPSEBTS-----EPSTLETCALIMKTIIC 1048  
 Db 886 GGDGLGVLILASCTPDLIQLNPTAKALAPYEGSLDIPVYKSEPPSPKSCDILIK--- 942  
 QY 1049 LEIYVYVKGSLDQSL-KDTL-----KFSIKRFPAYWSG-----YKSLAV 1088  
 Db 943 -DAYNNMLKHIDLAPRNSLIAGPAGSGHSLVFDLYASGNAYAEPPPIROGL 1001  
 QY 1089 -----HYAETEGSS-----CTSL-LEYQMLVSAM-----RMLIATTHADIMHL 1127  
 Db 1002 KETPLPSVGEVKGSLPVISVRKSSNSRSHYTIASALGSLNGLEKLFALILEPPSPLE 1061  
 QY 1128 TDSVVRQLFLDVL-----DGTAKLLVVPASVNCRLRGSKCTLLILRLQMKRE--- 1177  
 Db 1062 KLSKTPPTITIDSLKSKYKDYVITTSPIP-----LGSD-----LETFLOKKEGEGFIK 1110  
 QY 1178 -----IGSVDEILGLPTEILEGVLOADQOLMEKTKAVFSAFTVLOMKEMKSDIPQY 1231  
 Db 1111 LYSEGNIYDLDBRL-PLNLIBPAIVIOHTKVSPPKSSSLAISIVAPSLSEIWIYISOK 1169  
 QY 1232 SOLVIANVETIQOEVIALFDQTRSHLAGSATDEK-----DSMETDSCSRHRDQDGV 1286  
 Db 1170 KORKLSTYSLGWMKDKGRLYPEITHQLXSDHPEGRCLTCGGRGELIKISLEHKEKIAHY 1229  
 QY 1287 CVL---GLHLAKELCEVDEDEDSWLOVTRRL-----PILPTLLTTLLEVSLEMKONH 1335  
 Db 1230 TPLEFPLSFFPK-----SYMKEYQKLLKDBNASQPL--KLTLT-----KEFLN 1270  
 QY 1336 FTEATLH-----LLTLARTQO-----GATVAGAGITOSI-----CLPPL 1371  
 Db 1271 FCRSSSEFPKMMALMEQDLETESDPLIKPLALTSCPAKSGSLANDYANVRINTSL 1330  
 QY 1372 SVYQ-----ISTNGTQTPS-----AARSL-----DAPSPGVYR 1402  
 Db 1331 DIYEDATFLESFLNTIGTDTRSIIOQLMNRLLFTISKVGLSYTTLGQODRTLSDEGENYR 1390  
 QY 1403 LSNM-----LMEQLLKLRYNPLFEALDFGVHOERTLOCINAVRTQSLACL 1450  
 Db 1391 LHLAKKISINLTNIVLFEERPLSGHPODLPTIYQLL--KELVANNNTYIADIRSGSLI 1447  
 QY 1451 EEAADHY-----GFIILQSLNFM--KEWHFHLPOLM----- 1478  
 Db 1448 PHADHAIFLGPGSGPOGFLMDSPTCEVCSVDLHANVPQTEVCPKAPLISKANHTGSD 1507  
 QY 1479 RDIOVNIIGVLCQACTSLHSRKMLOHVLQNKNGGGLSVAQVRQRPSPA--ASAPSSS 1536

Db 1508 RTLKVNLSI-----HHION-----LKYSAFLHALVALVAGVSGS 1540  
 QY 1537 KOPADTEASBOALHTVQY-----LTKLSKTLAALRHFTPDVQIILDDSLDAEY 1590  
 Db 1541 GKTSLEBGFQKQBELLIKACTTTFPSDLVIDSHPIAS-----SQRSIDSTY 1587  
 QY 1591 NFLPALSFTTPTPDESEVAPSGTLLATVN--VALNMLGEL--DKKEPLTQAVGLSTQA 1645  
 Db 1588 -----FD--IAPSLRFAYSILTOAKALNISSTWPSWTWTKQCGSDCGGLGYQ- 1632  
 QY 1646 EGTRLKLSLMFTMENEFYLL-----ISQARY----- 1673  
 Db 1633 -----WIDRAFPALIEKRPCTCGFRIQPLAQEVLYEGKHFGELHTPIETVA 1680  
 QY 1674 LRDPVAPRDKQKQKQELASSELSTL-----LSLSIS-----RYPRGAPSSP 1714  
 Db 1681 LRPFPIKIQKP-LKALLDGLGYLPICQKLSLSVSEKTKALTAAYLYQTPETP 1734

RESULT 17  
 US-09-538-092-1154  
 ; Sequence 1154, Application US/09538092  
 ; Patent No. 6753314  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Glac, Iolc  
 ; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same  
 ; FILE REFERENCE: 15966-542  
 ; CURRENT APPLICATION NUMBER: US/09/538,092  
 ; PRIOR FILING DATE: 2000-03-29  
 ; PRIOR APPLICATION NUMBER: 60/127,352  
 ; PRIOR FILING DATE: 1999-04-01  
 ; PRIOR APPLICATION NUMBER: 60/178,965  
 ; PRIOR FILING DATE: 2000-02-01  
 ; NUMBER OF SEQ ID NOS: 1387  
 ; SOFTWARE: CuraPatSeqFormatter Version 0.9  
 ; SEQ ID NO 1154  
 ; LENGTH: 3210  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: misc.feature  
 ; LOCATION: (0)...(0)  
 ; OTHER INFORMATION: Polypeptide Accession Number P49454  
 ; US-09-538-092-1154

Query Match 1.4%; Score 129; DB 4; Length 3210;  
 Best Local Similarity 19.2%; Pred. No. 0.092;  
 Matches 361; Conservative 282; Mismatches 684; Indels 550; Gaps 99;  
 QY 18 RELMTLLGRSALRELSQLEALNKGRRLBELSLTYKPPSPSAAKRVANKDVASPLKE 77  
 Db 680 RNHNANVLDSSK-----VEVETQK-----LAYMELQKAEPSDQKHQKEI---EN 720  
 QY 78 LGLRISKFLGL--DEEQSVQL-----OCY--LOEYRGTRDSVK-----TVLDERQ 121  
 Db 721 MCKTSQQLGQVEDLEHKQLLSNEIMDKRCYQDILHAYESIRDLKSKDASIVYNEBH 780  
 QY 122 SQALLIKINDYYEERTCILRCVHLL--TYFODERHRYVE-----YADCVDK 168  
 Db 781 ORSLA-----FDQOPAMHNSFANIGQSGMPSERSSECRLEADQSPKNSAAILQVRVDS 834  
 QY 169 LE-----KEVSKYKQOFEELYKTEAPWETHNGVIMTEROVSRTV-----QCLRE- 214  
 Db 835 LERSLESQKQMSDLQKQCBELVQING--EIEENIMKAEQNHQSVFAETSGRISRLQBD 891  
 QY 215 -----OSMLLEIIFLYAYFEMAPSDLLVLT-IMFKRQG-FGSRQTNRLHLDVETMPFVDR 268  
 Db 892 TSHAGVVAETL-----SALENKEKEQLNDKVTETQAEIQGLKXSNHLEBSL-----KE 943  
 QY 269 IGYFSALLIVEGMDLESILKLCALDRRELDHQAQGLICQDMDCMLTFGDIPIHAPVLL 328

```

Db 944 LQLSETLSEKEMSIISL---NKREIELTQENGTLKEINA-----984
Qy 329 AMALLRHLPBETSSVVRKIGTAIQANVPOYLRLQLSLASGNDCTTSTACWVGL 388
Db 985 -----SLN-OEKNNLIQKSESFA-----NYIDEREKIS-1012
Qy 389 LSFVLTSLELHTLGNQOD-IIDTACEVL--ADPSLPBLFWGT-EPTSGIITLDSVCMF 444
Db 1013 -----ELSDQYQEKILILQRCBETGNAYEDLSQTKAQAQKNSKLECLINACTSLC 1064
Qy 445 PHLLSPILLQLRALVSGSTAKKYSFLDKMSFYNELYKHKPHDIVISHDGTLMRQTPK 504
Db 1065 ENRKHELQLEKAF-----AKEHQEPLTKLAFABE-----RNO---1097
Qy 505 LLYPLGGGTNLRIPGTYGVQW---LDDRATVLRREYSSTWTLPTCEILEMLHVVSTAD 561
Db 1098 -----NMLLETVQOALRSEMTDNQNNKSKEAGLKOELITLKEQNKQKEVND 1148
Qy 562 VIOHQVRVKPIIDLHVKYSTDLSDIADCLPITSITMLORLTVISPPVDIVASCVC 621
Db 1149 LLOEHEQLMKVKTGK-----EC-----QMLE---SEPI-----1174
Qy 622 LTVLAARNPAKVTWDLRH-TGFLPFVAHPVSSL-----QWISAQGNVAGGYMLMNS 674
Db 1175 -----RNSVKEBBERNQCNPKFQMDLEVKESILDSVNAQVQLEAM-----LNKK 1220
Qy 675 EOPGEYVTTAFLRLITLVKGGQSGTOSQGLVPCMFVKEMLPYHKRKYNSHGVRE 734
Db 1221 ELKQESKEKECLOHELOTRIGDLETSLNLDQMOQOESIGLMD-----1263
Qy 735 QIGCLILELHAILNCHETDLHSHTPSLOFLCISLAYTBAGOTVININGIVDTIDM 794
Db 1264 ---CEI-DAEKYISGPHELTSGQNDNAHQ---CSL-----QTMKMLN-ELKICE 1308
Qy 795 VMAAQPRSDGAE-GQGQGLIKITVLAFSV---TNVYIRLKPSS---NVVSP-841
Db 1309 ILQAEKYELVTELNDRSRECTATRKMAEYGVKILNBYKILINDDGLHGLVEDI PGGE 1368
Qy 842 --EQALSQHGAGNNLIVLAKYITHK---DPLPLRLAIOQLRLATVAPMSYACLG 895
Db 1369 FGEOPNEHPVS---LAPLESNSYEHILTSDKQVOMHPALQKPLSTL-QSHKIL-1421
Qy 896 NDAARIPARFLRLQSKIEDMRKIMLEFLTVAVETOPGLLEPLNTEVDKSGSGSKF 955
Db 1422 HDQHCQSSKXSELOTYVDSLKAENLVLS-----TNLRNFG-DLVKEM 1464
Qy 956 SLGMSCLHAYLELIDSOQODRYWCP---LHRAAIAFLHMLQDRDRSAMVLRTKP 1011
Db 1465 QLGLEBGL-VPSLASS-----CVPSSSLSSLGDSFYALLQGTGMSL-----1509
Qy 1012 KFWENLTSPPLFGTSSPSETSEPSILETCALIMKIIIC-LEIYVYVKGSLDQSLKOTLK 1069
Db 1510 -----SNLEGAVS-----ANQCS-VDEVFCSLQTYVDSLKAENLVLSNLRN 1551
Qy 1070 FS-----IKRFAVMSGYKSLAVHVAETEGSSCTSLLEVOMVSARMMLIIATHADIM 1125
Db 1552 FQGLVKNEMQGLEGLVPSLSSCV-PDSSSLSLG---SSFYRLRLBGTGMSLIS 1606
Qy 1126 HLTDSVVRQFLDVLDTGKALLLVASVNCILRSGMKCTLLILLRLQMKELSGVDEIL 1185
Db 1607 NLEGVVSNQCSVD-----EYFCSSLOEENIL-----RKEETPAP---1641
Qy 1186 GPLTEILEGVLOADQOLMEKTKAKYFSAPITVLOKEMKVSIDIPOYSQVLNVCTLOE 1245
Db 1642 AKGYEELSELCBVYFQSLKEKESKESOGI---MKNKEIOELBOLILSSERQELCLKQ 1697
Qy 1246 VIALFDOTRHSIALGATEDKDS-----METDQSRSRH---RDQRDV-----CVLGH 1292
Db 1698 YLSENEQMOQKLT--SVTLMESSKLAAEKQTEQSLLEVARLLOGLDSSRSLGLGD 1755
Qy 1393 L-----AKELCEVEDDEDSDMLQVTRRLPILFTLLTLEVSLRMKNLH---FTBATLH 1343
Db 1756 TEDAIQGNESCDISKENTS--ETTERTP-----KHVHQICDQKQAOQL 1798

```

```

Qy 1344 LITLAR--TQAGATVAVAGITQSTCLPLLSVYQSTNGTNGTQAPSPASRSKLAPSWGVYR 1402
Db 1799 NDIETKITETGAVKPTGE-----C-----SGEOSPDVTNYP-----PGEDK 1834
Qy 1403 LMSIMEQLLKTRLRYNPLPEAL--DFVG-----VA--OERTLOCLINAVTQSLACLBEA 1453
Db 1835 TQGS--SFCISELSFSG--PNMLVPMDFLQNGEDIHNLQLRKVENSN--ENRRLHVIDR 1889
Qy 1454 DHTVGFILQSNFKE-----WHFILPOLMRDIQVNLGYLCOACTSLSHSRKMLQHYLQNK 1509
Db 1890 DRKVESLL--NEMKEILDSKHLQEVQMLTKI-----BACIEL-----EKIVGELKKE 1934
Qy 1510 NGDGLP-----SAVQORVQRPSPAS-----AAPSSQOPADTFA-----SQOALH 1552
Db 1935 NSDISEKLEYFSCDHQELQEVETSEGINSDLEMHADRSSREDGDVNAKVNSWKEPRL 1994
Qy 1553 TVQYGLTKILSKTLAALRHFTPDVQCIILDOSLDLAEYNPLFALSFTTPTPDS-----VA 1608
Db 1995 DVENELSRIRSEK-ASIHBA-----LYLEADLBYVQTEKLCLEK-----DNEKKQYI 2042
Qy 1609 PSFGTLATVVALNML-GEIDKXKEPLTQAVGLSTQ-ABGTRILKSILMFTMENC FYLL 1666
Db 2043 VLEBELSVTISRQQLRGEIDTMSKTTALDQLSKKKEKETOELSHQ-----2091
Qy 1667 ISQMRVLRDPAVHRDQKQKQELSSSLSTLSSLSYFRGAPSSPATGVLPSPQGS 1726
Db 2092 -SELCICIQVAEYKETEILLQTLSDVSEILKDKTHLOEK-----LOSLEKDS 2140
Qy 1727 TSLKASPESEPLIQL 1743
Db 2141 QALSTKCELENQIQL 2157

RESULT 18
US-08-353-700-1
; Sequence 1, Application us/08353700
; Patent No. 5599919
;
; GENERAL INFORMATION:
; APPLICANT: YEN, TIMOTHY J.
; APPLICANT: RATNER, JEROME B.
; TITLE OF INVENTION: NUCLEIC ACID ENCODING A
; TITLE OF INVENTION: TRANSLIENTLY-EXPRESSED KINETOCHORE PROTEIN,
; TITLE OF INVENTION: AND METHODS OF USE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DANN, DOREMAN, HERRELL AND SKILLMAN
; STREET: 1601 MARKET STREET, SUITE 720
; CITY: PHILADELPHIA
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2307
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/353,700
; FILING DATE: 09-DEC-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: REED, JANET E.
; REGISTRATION NUMBER: 36,252
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 563-4100
; TELEFAX: (215) 563-4044
; INFORMATION FOR SEQ ID NO. 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3248 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear

```

MOLECULE TYPE: Protein  
 HYPOTHETICAL: NO  
 ANTI-SENSE: NO  
 ORIGINAL SOURCE:  
 ORGANISM: HUMAN  
 US-08-353-700-1

Query Match 1.4%; Score 129; DB 1; Length 3248;  
 Best Local Similarity 19.2%; Pred. No. 0.094;  
 Matches 361; Conservative 282; Mismatches 684; Indels 550; Gaps 99;

18 BELWTLIGRSALRELSQIEALINGMRLLLEGLSYRPPSPSSAEKVAANDVASPIKE 77  
 680 RULHNVLDSK-----VEVETOK-----LAVMELQQAKESSDQKHKEI-----EN 720  
 78 LGLRIKSKLGL--DEEGSVOLL-----QCY--LQEDYKTRDSVK---TVLQDERQ 121  
 721 MCLKTSQLTGVQVEDLEHKLQLLSNEMDKRCYQDLHAEYSILRDILSKSDASLVTNEDH 780  
 122 SQALLIKIADYYEERTCILRCVLHL--TYFQDERHPRYVE-----YADCYDK 168  
 781 QRSLLA-----FDQPRAMHNSFANIIIGQSGMPSEBSECHLEADQSPKNSAIIQNRVDS 834  
 169 LE-----KEVSKYRQGFEEIYKTEADPTWETHGNLMTEROVSRMFV-----QCLRE 214  
 835 LEFSLSEOKOMNSDLQKOCSEELVQIKG--EIEENIMKAEQWHQSGFVAFETSORISKLED 891  
 215 -----QSMLEIIFLVYAFEMAPSDLVLT--KMFKEQG--FGSRQINRLINDETMDFPDR 268  
 882 TSAHQNVAAETL---SALENKEKELQLLNDKVETEQAIEQLKKSNNLLRDSL---KE 943  
 269 IGYFSALIVGMDIESLHKCALDRREIHOFAODGLICODMDCIMLTFGDIPIHNAVLT 328  
 944 LQLSETSLTEKEMSSIISL--NKREIBELTQENGTLKEIN----- 984  
 329 AMALLRHILNPEEISSVVRKIGGTAIQNLNVQYITRLIQLSASGNDCTSTACMCYGL 388  
 985 -----SLN--QEKNNLQKSESFA-----NYDEREKSSIS----- 1012  
 389 LSFVLTSLLEHLTIGNOOD-IIDTACEVL--ADPSLPFLFMGT--EPTSGIILIDSVGMF 444  
 1013 -----ELSDQYKQELIILQRCSEETGAYEDLSQYKAAQEKNSLLECLNECTSLC 1064  
 445 PHLSPLQLLRALVSGKSTAKKAYVSLDKMSFYNELYKHPHDVISHEDGLMRQTPK 504  
 1065 ENRKNELEQLKEAF-----AKEHQEFLTKLAFAE-----RNG--- 1097  
 505 LLYFLGQGTNLRIGQTVGVGM---LDDRATLVMEWYSYSWTITFCEIEMLHVASTAD 561  
 1098 -----NMLELETYQALRSEMTDNQNSKSEAGLKOELMTLKEBNKMKQKEVND 1148  
 562 VIQHCRVKPIIIDLVHKVISTDLISADCLAPITSRIVMLQRLTTVISPVDVASCVC 621  
 1149 LLQENEGOLMKMKTKH-----EC-----QNLN---SEPI----- 1174  
 622 LTVLAANRPKAVTWDLRH--TGFLRPVAHPVSSLS-----QMSABEKNAGGYGNLLMS 674  
 1175 -----SNSVYERESERNQCNFKPOMDLVEKEISLDSYNAQVLTQLEAM-----LRNK 1220  
 675 EOPGEGVTVIAPFLRLITLVKQGLSTQSGQLPVCWPFVLYKMLRPYHKRYNSHVR 734  
 1221 ELKQISEKEKECQHELOTRGLDLETNSLDQMSQETISGLKD----- 1263  
 735 QIGGLIELIHAIINLCHETDLHSHRPSLOFLCISLAYTEAQTVINIGICVDTIDM 794  
 1264 ---CEI--DAEBKYISGHEHLSIQNDNAHLQ---CSL-----QTMNKLN--ELEKICE 1308  
 795 VMAAPRSDAE--GQGGQGLIKTVKLAFSV---TNVNIKPPS-----NVVSPL--- 841  
 1309 ILQEKYELVTELNDSSSECTATARKAAEEVGKLLNCKTLNDSSGLHSELVEDIPGE 1368  
 842 --EQALSQGHAGNNLLAVLAKYIYKH---DPALRLAIQLLKLRLATVAPSVYACIG 895

1369 FGEQPNQHPVS-----LAPIDESNSYHLLTSDKEVQMHFAELOEKLST--QSEHKIL- 1421  
 896 NDAAIRDAFLTRLOSKIEMDRKIMLIEFLTVAVETQPGILIELFLNLEVDGSDGSEK 955  
 1422 HDQHCQMSKMSSELQTVVDSLKAENLVLS-----TLNRNQG--DLYKEM 1464  
 956 SLGWSCLHAYLELIDGQODRYGCP---LHRAAIAPLHLMQDRDSAMVLATKP 1011  
 1465 QLGLEBGL--VPSLSS-----CVPDSSSLSGDSSFYRALLEOTGDMSSL----- 1509  
 1012 KFMENLSPLEGLTSPSETSSEPILETALIMIKIC--LEIYVVGSLDQSLKDTLTKK 1069  
 1510 -----SNLEGANS-----ANQCS--VDEVFSSIQTYVDSLKAENLVSTLKN 1551  
 1070 FS-----IEKPAWSGYKSLAVHVAETEGSSCTSLLEYQMLVSAWMLLIATTHADM 1125  
 1552 FQGDVYEMOQLGLEEGVLPSSSCV--PDSSSLSLD-----SSFYRALLEOTGDMELLS 1606  
 1126 HLTDSVVRQLFLVDLGTKALLVPAVNCVLGSKCTLLILLRQMKRELGSVDEIL 1185  
 1607 NLEGVSNAGQSYD-----EVFCSSIQEENLT-----RKETPSAP-- 1641  
 1186 GPLTEILEGVLOADQOLMEKTKAKVSAFTVLQMKEMKYSDIPOYSQVLVANCETLOE 1245  
 1642 AKGVBELESICEVYRQGLEKLEKMSQGI---MKMKELGLEQLLSSRQELDCLRQ 1697  
 1246 VIALFDQTRSLALGSTEKDS-----METDDCSNRH--RDQRDGV-----CVLGLH 1292  
 1698 YLSENEGMOQKLT--SYTLEMESKLAKEKQOTQLSLEFVARLOGLDLSRSRLGID 1755  
 1293 L-----AKELCEVDEDEGSMLOVTRRLPIPLTLTLTLEVLMMKONLH--PTEATLHL 1343  
 1756 TEDALQARNNSCDSIKENHS--ETTERTP-----KADVHQCCKDAQOQL 1798  
 1344 LTLTAR--TQCGATAVAGAGITQSICLPLSLVQSLNGTAGTPSASRKSIDABSPGVYR 1402  
 1799 NLDIEKITERGAVKPTGE-----C-----SGRQSPDTYVER-----PGEDEK 1834  
 1403 LMSIMBQILKTLRYNPLREAL--DEVG-----VH--QERTLOCNAVRVQSLACLEEA 1453  
 1835 TQGS--SECSISELFSFG--PVALVPMDFLQNGEDIHNLQIRAKVETSN--ENLRLLHYIEDR 1889  
 1454 DHTVGFILQLSNEMKE-----MHFRLPOLMRDIOVNLGVCOACTSLSHSKMLQHYLQNK 1509  
 1890 DRKYSLSL--NMKELDSKLHGEVQMTKI-----EACIEL-----EKVIGELKE 1934  
 1510 NQDGLP-----SAVADQVRPPSAAS---AAPSSSKOPADTEA---SEQOALH 1552  
 1935 NSDLSEKLEBYFSCDHQELLQRYVETSEGLNSDLEMHADKSSREDIGDNVAKVNDSWKERFL 1994  
 1553 TVQYGLKTLISKTLALRHFTPDVCOILLDQSLDLAEVNFPLASTTTPPSE-----VA 1608  
 1995 DVENEISRLRSEK--ASIEHBA---LYLEADLEVVQTEKLEK-----DNEKQKAT 2042  
 1609 PSFGTLATVNAVNLNL--GELDKKEPLTQAVGLSTQ--AEGTTLKSLMFTWENCFYLL 1666  
 2043 VCLBEELSVYTSERNQRLRGELDTMSKKTALDOLSKMKKEKTOBELSHQ----- 2091  
 1667 ISQAMRYLDRPAVRHPRDKQMKQELSELSTLSSLSRFRKGAAPSAPATGVLPSPQGS 1726  
 2092 --SECLHCIOVAERAEVKEKTELLQTLSSDVSELLKDKTHLOEK-----LOSLKDS 2140  
 1727 TSLSKAPESQEPILQL 1743  
 2141 QALSLTKCELBNQIAQL 2157

RESULT 19  
 PCT-US95-16216-1  
 ; Sequence 1, Application PC/TUS9516216  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Yen, Timothy J.  
 ; APPLICANT: Raltner, Jerome B.

```

: TITLE OF INVENTION: Nucleic Acid Encoding a Transiently
: TITLE OF INVENTION: Expressed Kinetochore Protein, and Methods of Use
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Dann, Dorfman, Herrell and Skillman
: STREET: 1601 Market Street Suite 720
: CITY: Philadelphia
: STATE: PA
: COUNTRY: USA
: ZIP: 19103-2307
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US95/16216
: FILING DATE:
: CLASSIFICATION:
: PRIORITY APPLICATION DATA:
: APPLICATION NUMBER: US 08/353,700
: FILING DATE: 09-DEC-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: Reed, Janet E.
: REGISTRATION NUMBER: 36,252
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (215) 563-4100
: TELEFAX: (215) 563-4044
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3248 amino acids
: TYPE: amino acid
: STRANDEDNESS: not relevant
: TOPOLOGY: not relevant
: MOLECULE TYPE: protein
: HYPOTHEICAL: NO
: ANTI-SENSE: NO
: PCT-US95-16216-1

Query Match 1.4%; Score 129; DB 5; Length 3248;
Beet Local Similarity 19.2%; Pred. No. 0.094;
Matches 361; Conservative 282; Mismatches 684; Indels 550; Gaps 99;

QY 18 RELMTLLGRSLRLSGLAEALNGWRRLLEGLSYVPPSPSSAEKTKANDVASPLKE 77
DB 680 RNLHNVLDSSK-----VEVETQK-----LAYMELQQAEPSSDQKHQKEI-----EN 720
QY 78 LGLRISKEFLGL--DEEGSVQL-----QCY--LOEDYRGTDRSVK---TVLQDRQ 121
DB 721 MCLKTSQLTGQVEDLEHKLQLLSNIMDKRCYQDLHAEYESLRDLKSKASLVTNEDH 780
QY 122 SQALLIKLADYYEERTCILRCVHL--TYFODERHPYVE-----YADCYDK 168
DB 781 QRSLLA-----FDQOPAHGHSFANITIGQSGMPSESRCSRLDADQSPKNSAILQNRVDS 834
QY 169 LE-----KELVSKRQGFEEELYKTEAPTWETHGMLTERQVSRKFV-----QCLAE- 214
DB 835 LEFSLESQKQMSDLOKQCEBELVQIKG---ETIENMLKRAEQHQSFVAETSORISKQED 891
QY 215 -----QSMLEITFLYVAFEMAPSDLVLT--KMFKEQG--FGSQRNRLVDETMDPFYDR 268
DB 892 TSAHNVAVAEVL-----SALENKEKELQILNDKVTETQAEIQGLKKSNNHLESL---KE 943
QY 269 IGYFSAILLVEGMDESILHKKALDRRELHQPADQGLICQDMDCMLTFGDIPIHAPVLL 328
DB 944 LQLLSETLSLEKKESSITSL---NKRRIEELQENGTLKEINA----- 984
QY 329 AVALLRHLNPEBETSSVVRKIGTAIQANVFOYLTRLQSLASGNDCTTSTACMCYGL 388
DB 985 -----SLN--QEKNNLIQKSESFA-----NYIDEREKISIS----- 1012
QY 389 LSFVLTSLEHLTIGNQOD-IIDTACEVL--ADPSLPFLFWGT-EFTSGIGITLDSVCMF 444

```

```

DB 1013 -----ELSDOYKQEKILLQRCCEGTGNAYEDLSQKYKAQEKNSKLECLNECTSLC 1064
QY 445 PHLSPLQLLRALVSGSKTAKYVSPIDKMSFNLKYKHPRHVISHEDTLMRRQPK 504
DB 1065 ENRRKELEQLKEAF-----AKEHQEFLTKLAFABE-----RNO--- 1097
QY 505 LLYPLGGQTNLRIPQGTVGQVM---LDDRAVLVMEVSYSSWTLFTCEIEMLLHVVSTAD 561
DB 1098 -----NLMLELFTVQALRSEMTDNNKNSSEAGLKOELMTLKEQNKQKEVND 1148
QY 562 VIOHQRYKPIIDLVHKVISTDLISADCLPITSRIYMLQLRTLVTSPPVDVIAASCVC 621
DB 1149 LLQENEGQMKVMYKTKH-----EC-----QNLB---SEPI----- 1174
QY 622 LTVLAARPAKVTWDLR--TGFLLPVVAHPVSLS-----QMAEGNAGAGYGNLMS 674
DB 1175 -----RNSVKEKRSERQCNFKPMQMDLEVEYESIDSYNAOLVQLEAM-----LENK 1220
QY 675 EOPQGEYGVTLAFLRLITLVKQGLSGTQSGLVPCVMFVLKEMLPYHKWRVNSHVRE 734
DB 1221 ELKIQSEKEKECELOHELTQTRGDLTSSNLDQMOSQELSGKD----- 1263
QY 735 QIGCLILELHAIINLCHETDLHSSHTPSLOFLCISLAYTEAGQTVININGIGVDTIDM 794
DB 1264 ---CEI--DAERKYISGPHELSTQNDNAHLQ-----CSL-----QTWNKLN--ELEKICE 1308
QY 795 VMAAQPRSDGAE--GQGQGLIKTYKLAFSV---TNVYIRLKPFS-----NVVSPFL--- 841
DB 1309 ILOAEKVELVTELDNRSECTIATRKAAEYGVKLNEKYLINDSGLHAGELVIEDIPGE 1368
QY 842 ---EOLASQHGAGNNLAVLAKYIYHKH---DPALPRLAIQLRLATVAPMSVYACLG 895
DB 1369 FGEQPNQHPVS---LAPLDESNYELHTLSDKEVQNHFPLEQKPLSL--QSEHKTL- 1421
QY 896 NDAAIRDAPFLTRLQSKIEDMRIKYMLIEFLTVAVETOPGLIEFLNLEVQSGDSGEF 955
DB 1422 HDQHQMSKMSSELQTYVDSLKAENLVLS-----TNLRNFG--DLVXEM 1464
QY 956 SLGMSCLHAYLELIDSGQODRYWCP---LHRAAIAFLHAIQMODRRDSAMVLRKTP 1011
DB 1465 QLGLEBEGI--VPSLSSS-----CVPDSLSLSLSDGSPFRALLEQTDMSL----- 1509
QY 1012 KFMENLTPLFGTSPSESTSEPSILETCALIMKIC--LEIYVYVKSLSQSLKDTLKK 1069
DB 1510 -----SNLEGAVS-----ANQCS--VDEVFCSLQTYVDSLKAENLVSTLNRN 1551
QY 1070 FS-----IEKRFAYVSGYKSLAVHVAETEGSSCTSLLEYQMLVSAWMLIITATTHADIM 1125
DB 1552 FQGDLYVEMQGLEBGLVPSLSSCV--PDSSLSLSIGD---SSEFYRALLQOTGDMSLSS 1606
QY 1126 HLTDSVVRQQLFLDVLDTKALLLVPAVSNCLRLGSMKCTYLLILLRWKRELGSVDEIL 1185
DB 1607 NLEGVASAMQCSVD-----EVFCSSLOEBNLT-----RKETPSAP--- 1641
QY 1186 GPLTEILLEGVQADQOQMLEKTKAVVSAFIVLMKEMKVASDIPQYQSLVANTVCELTQEE 1245
DB 1642 AKGYVEELSLCEVTRQSLKEKEMESQGI---MKKEIODELELSSSEQELDCLRKQ 1697
QY 1246 VIALFDQTRHSLALGSAATEDKDS-----METDCCSRRH--RDQRDGV-----CVLGLH 1292
DB 1698 YLSNEHQOQGLT--SVTLMESSKLAAKKQTEQLSLELVARLQLOGLDLSRSSLIGID 1755
QY 1293 L-----AKEICEVDEDEGSMLOQYTRRLPIPLTLITLVLSLRMKQNIH---FTBATLHL 1343
DB 1756 TEDAIQGNRESCDISKETKTS--ETTERTP-----KHDVHQICDDDAQODL 1798
QY 1344 LTLTLAR--TOGATAVAGAGITQSICLPLLSYVQUSTNGTAQTPASRSKSLDAPSPGYVR 1402
DB 1799 NLDIEKITEGAVYPTGE-----C-----SGEOSPOTNIER-----PGEDK 1834
QY 1403 LMSIMEQLLKTURLYNNFLPEAL---DFVG-----VH--QERTLOCLANAVRTQVSLACEEA 1453
DB 1835 TQGS--SECSISELSPSG--PNALVPMDFLGNQEDIHNLQRLVKEKSN--ENLRLLHVLIEDR 1889

```

QY 1454 DHTVGEIILQSNFMKE-----MHFHLPEQLMRDIOVNLGYLCOACTSLHSRKMLOHYLQNK 1509  
 DB 1890 DKRVESL---NEMELDSKHLQSVQMTKI-----EKICEL---EKIVBELKKE 1934  
 QY 1510 NCGDLP-----SAVQRYQRPSSAAS---AAPSSSKOPADTEA---SEQOALH 1552  
 DB 1935 NSDLSEKLEKFEFSCDQHELLQRYVETSEGLNSDLEMHADKSSREDIDGNVAKVDSKKEFPL 1994  
 QY 1553 TYOYGLIKLSTKTLAALHFTPDVCOIILDDOSDLAEVNFLEFALSTFTPTDSE---VA 1608  
 DB 1995 DVNELSRLRSRK-ASIEHEA-----LYLEADLEVQTEKCLEK-----DNEKKQKVI 2042  
 QY 1609 PSFGTLIAVNVANLML-GELDKKEPLTQAVGLSTQ-AEGTRTKSLMFTMENCYILL 1666  
 DB 2043 VLEELSELVYTERNQRLGELDTMSKKTALDQLEKMEKTOLESHQ-----2091  
 QY 1667 ISQAMRYLDPVHPDKQRMKQELSELSTLSSLSRYFRGAPSSPATGYLPSPQGS 1726  
 DB 2092 -SECHICIQVAAEAYEKEKTELLQTLSDVSELKDKTHLOEK-----LQGLEKDS 2140  
 QY 1727 TSLSKASPESSOEPLQL 1743  
 DB 2141 QALSLTKCELENOIAQL 2157

RESULT 20  
 PCT-US94-00198-3  
 ; Sequence 3. Application PC/TUS9400198  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Schering Corp.  
 ; TITLE OF INVENTION: RAS Associated GAP Proteins  
 ; NUMBER OF SEQUENCES: 6  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Schering Corp.  
 ; STREET: 1 Giraldo Farms  
 ; CITY: Madison  
 ; STATE: New Jersey  
 ; COUNTRY: USA  
 ; ZIP: 94304-1104  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: Macintosh  
 ; OPERATING SYSTEM: 6.0.8  
 ; SOFTWARE: Microsoft Word 5.1a  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: PCT/US94/00198  
 ; FILING DATE:  
 ; CLASSIFICATION:  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/004,824  
 ; FILING DATE: 15-JAN-1993  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Lum, Paul G.  
 ; REGISTRATION NUMBER: 32,743  
 ; REFERENCE/DOCKET NUMBER: DX0352 PCT  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: (201)822-7255  
 ; TELEFAX: (201)822-7039  
 ; INFORMATION FOR SEQ ID NO: 3:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 2938 amino acids  
 ; TYPE: amino acid  
 ; STRANDEDNESS: single  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; ORIGINAL SOURCE:  
 ; ORGANISM: Saccharomyces cerevisiae  
 ; PCT-US94-00198-3

Query Match 1.4%; Score 128.5; DB 5; Length 2938;  
 Best Local Similarity 19.2%; Pred. No. 0.088;  
 Matches 183; Conservative 149; Mismatches 351; Indels 271; Gaps 41;

QY 971 DSQOQDRYWCSPILHRAI-AFLHALMQDRDSAMLVLRKPKKMENTSPFGLSPPS 1029  
 DB 807 ERSQSPSSPPQOQSDLPSPSVSSAGFSNHSITATPTLLKNKSP-----KPN 860  
 QY 1030 ET-----SEPSILETCALIMKICLEIYVYVKSILQSLQDKTLKKKSIEKRFAYW- 1079  
 DB 861 KTKIADKQKQKQSYSR-----VILSDNBARKIMNITSIFKGMNWF 905  
 QY 1080 -----SGYKSLAVHVAETEGSSCTSLLEYQMLVSNMRMLIIATTHADIMH 1126  
 DB 906 IRPDANTEPKTFPDILKPLFVSIIDS-----NQRLOVTAFAFIEIPLSYIATFED 956  
 QY 1127 LTVSVAROLFVDLTOSTKALLVPASVNCRLRGSMKCTLLILLRQMKRELSYDEITG 1186  
 DB 957 IDNDLDFRVLNDHYLCTTAVVTLFASLPDKLENARREMLDIIVKFOR---VRSYLS 1012  
 QY 1187 PLTE---ILEGVLAQDQMLEKTKAKVFA-----FTVLOMK 1221  
 DB 1013 NLAKEHNLVQAIIITTELTLPVLVAVGSGIFISYCSRGNTRLIKISCEFLRSIRFY 1072  
 QY 1222 EMKYSDIPOYSQVLVNCETLQEE-----VIALPDQTRHSIAL--GSATEDKDSMET- 1271  
 DB 1073 QKYVGALDOYSIYNIDPIDAMAQDNFTAGSVALORLRNIIITLYIKGSDSILDSMDVI 1132  
 QY 1272 -----DDGSRHRDQ-----RDGVCVLGHLAKELCEVEDGDSMQLQVTR 1312  
 DB 1133 YKMWFFYSCKSKSTQOEVLDFRSLAGILASMSGILSDMOLEKSKAPDNDEGS-LSFES 1191  
 QY 1313 RLPLPLTLTLTEVLSLRMKONLHFTATYHL---LTLARTQOGATAVAGAGITQSICL 1368  
 DB 1192 RNPAV-EVHKSLKLELTKKKNFPISSQOCWLNPNLLTHRENSD-----IISTEL 1240  
 QY 1369 PLISVQOLSTN-----GTAQTPSASRSLDAPBPVGYRLSMSLMQLEKTLR----- 1416  
 DB 1241 HPLSPMLFNNLGLKIDELMSIDLSKSHDS-----SFVLBOIILIIIFTLIKRSD 1291  
 QY 1417 -----YNFLPEALDFVGHQERTLOCLNNAVTVQSLAEADHTGFIQLSNFMKEW- 1470  
 DB 1292 DEKIMLLFSTDLDVAD-----KLEIYEKRS-IKSKRYYG-IIQNSKMPRAFE 1339  
 QY 1471 -----HFLPLQMR-----DIOVNLGYLCOACTSLHSRKMLOHYLQNKDGD 1513  
 DB 1340 HSEKNLGISNHFFLKKKMKLVIGWFKLSINKDYDENL---SPLREMDLQKDEDF 1394  
 QY 1514 L-----PSAAGQVQRPSSAASAPSSSKOPADTEASEQALHTYQY-----LTK 1560  
 DB 1395 LYIDTIESAKALAYITHNP--LEIPSSSKK--DMNRSS-----TVSGNHFTILK 1444  
 QY 1561 ILSKTL-----AALRHFTPDVCOIILDDSLDAEYNFLFALSFTTPTDSEVAPSGTUL 1614  
 DB 1445 GLEKSAIDNQFPVSLHKISLINENVIATLNLNSNANVAVSLKFTLP---MGSPKQDR 1501  
 QY 1615 LATVNVANLML-----GELDKK-----KEPLTQAVGLSTQAEGRTRTKS- 1653  
 DB 1502 IAFRLAFIDIVTVYPVNPPEKHEMDKMLAIDFLKYLIIKNPIIAFFSGLACSPADVLDVAG 1561  
 QY 1654 ---LIMFTMENCFLILISQAMRYLDPVHPDKQRMKQELSEL---LSTLSSLSRYF 1706  
 DB 1562 GFINAPDTRASHITVTELL-----KQETKRAARDDIIRNSCATRALSLVT 1609  
 QY 1707 RRGAPSSPATGYLPSPQ-----KSTSLSKASP--ESQEPILQVQAFVRRHMQR 1753  
 DB 1610 RSRGNKYLKTLRPVLQGIVDNKSEFEDIKMFGENSEKMDL---FEKVMTR 1660

RESULT 21  
 US-09-538-092-1136  
 ; Sequence 1136, Application US/09538092  
 ; Patent No. 6753314  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Glot, Loic  
 ; APPLICANT: Mansfield, Traci A.

TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same  
FILE REFERENCE: 15966-542  
CURRENT APPLICATION NUMBER: US/09/538,092  
CURRENT FILING DATE: 2000-03-29  
PRIOR APPLICATION NUMBER: 60/127,352  
PRIOR FILING DATE: 1999-04-01  
PRIOR APPLICATION NUMBER: 60/178,965  
PRIOR FILING DATE: 2000-02-01  
NUMBER OF SEQ ID NOS: 1387  
SOFTWARE: CurePathSeqFormatter Version 0.9  
SEQ ID NO 1136  
LENGTH: 3433  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (0)...(0)  
OTHER INFORMATION: Polypeptide Accession Number P46939  
US-09-538-092-1136

Query Match 1.4%; Score 127.5; DB 4; Length 3433;  
Best Local Similarity 18.6%; Pred. No. 0.15;  
Matches 389; Conservative 256; Mismatches 695; Indels 749; Gaps 95;

QY 8 TSVLSFCSRSREIWTLLGRSALRELQIAE-----LTKHMRLLBGLSYKKPPSPSA 62  
DB 364 SSVGVSLDAGNOLIT---QGTLSDEEFETIOEOMTLNANREAL-----RVESM 409  
QY 63 EKVANKDVASPLKELG-RISKPLGLDEEOSVOLQCYLOEDYKRTDSVKTVLQDERQ 121  
DB 410 DROSLHVLWMLQKQQLQSLAWLTLTEBRLOKMETCPDDVY---SLQKLEHKHS 465  
QY 122 SQALILKADYYBERTILKCVLHLTYFODERHPYVEVADCYDKLEKLSVKRYQQF 181  
DB 466 LQSDI-----EAEQVKVNSLTTHVVIIVDENG---ESATAI--LEDQI-----QKL 506  
QY 182 EELYTAEPATWETHNLMTERQVSRMFVQCRLREOSMLEITFLVAYAFEMAPSDLLVLTK 241  
DB 507 GERW-TAVCRW-----TE--ERW-----NRLQELINIMQELLE--EQCLLKWL 545  
QY 242 MFEQGFSGSRQTNRLVDETMDPFVDRIQYSALILVEGMDIESLHKCALDRRELHQA 301  
DB 546 TEKEBALNKVQTSNFKQKELSVSRLLA-----ILKEDMEK-----RQTLQQLS 591  
QY 302 QDGLICQMDCLMTFFDIPHHAPVLLAMALRLHNLNEERTSSVVRKIGTGAIQLNVQY 361  
DB 592 EIG---QVVGQL-----:-----DNSKASKINSSEBLT--QR 620  
QY 362 LTRLQLSLASGNDCTSTACWCYGLSPVLTSLHLHTIG---NODIIDT-----A 411  
DB 621 WDSLQRLSDSSNOYQAVNA-----KLGMSQIPQKDLLETAVAREQA 662  
QY 412 CEVLADPSLPFLFWGTEPTSGILILD-SVCGMPHLSPLQL-LRLALVSGSKTAKKY 469  
DB 663 ITKSKSKQLRP---PPPKKQRIHVDIEAKKKPAISABELLNLKMKTAIQTEIKEX 718  
QY 470 SFLDMKSYNNELYKRPKPHDVISHEDDTMRQTPKLLYPLGGQINLRIPQ---GTVCQV 525  
DB 719 M---KMOQTSW-----KKKLKALEKQCRERIPRADELNQTQOI 754  
QY 526 MLDDRAVILVRMEYSWTLPTCEIEMLLHVST--ADVIOHCQVKRIILD----- 575  
DB 755 LVE-----QMGKEGLPTEIKVLEKVSSEKKNVSOHLEDEKRIQOEDINAYF 804  
QY 576 -----VHKVIST-----DLSIADCLPITSRIYMLQR-LTTVS--PPYDVI-ASGVN 620  
DB 805 KQLDELKVKITKEKEMVGHSTISSESSRQSLPSLKXSCQRELNLNLGLPKIEMARASOSA 864  
QY 621 CLTVLAANPAKVMTDLRHGTFLP-----VAHPVSSLSQMTISAEKMANAGGYGNL----- 670  
DB 865 LMSQPSA-----PDPVORGFDSPFLAGRYQAVQVQVVEDROOHLLENELKGPQHALETFLK 917  
QY 671 -----LMNSEQPGEGVGTIAFLRLITTLVKGQLSGTQSGVLPVCMFVLKEMLPYHKMR 726

DB 918 TLKVLNDSENKAQVS-----LVNLDLAVKEXALQSKTLLDE-----ILENQKALMKLA 968  
QY 727 YNSHGREBQIGCLILELIIHA-----ILNCHETDLHSHTPSLQFLCISLAYTE 776  
DB 969 EETKALEKKNVHPDEKLYKQEFDDVQKWNKLKVLVSKDILLLEIAL-----TLRAFE 1022  
QY 777 AGQTVINIMIGV-DTIDMVAAPRSDGAEQ-QGQQLIKTVKLAFSVTNNV----- 828  
DB 1023 ADSTVIEKMDGVDFLMKQQAAGDDAGLQRLQDQSAFANEIETLESSLKNKKEIETN 1082  
QY 829 IRLKPSNVVSPLEQALSOHGAHGNL---IAVLAKTYIYHNDPALPRLAIQLIKRLATV 885  
DB 1083 LRSGPVAGIKTWQTRIGDYOTOLEKSKELATQKSLSSQEK-----AANLKQDLAEM 1137  
QY 886 APMSVACLDNDAADAPLTR-----LQSKIEDM-----RIKWMILE 924  
DB 1138 QEMWTOA-----BEEYLERDFEYKSPEELBSAVEEMKRAKEDVLOKEVRVKILKN 1188  
QY 925 FLTVAVETQPLLEFLNLBVK-----DGSQSKFSLGMWSCILHAVLELDSQQOD 976  
DB 1189 IKLIAAVPSSGQELTSELVNLVLENYQLCNIRGKCHTLEBWSCWIELHLHYD----- 1243  
QY 977 RYWCPLIHRAAIAFLHALMODRDSAMLYLRTPKFEMNLSPFGLTSPSESEPSI 1036  
DB 1244 -----LETWMLTEBRMKSTEVLPKTDVANEAL-----ESLESVL 1280  
QY 1037 LETCALIMKIICTEIVYVVGSLDOSLKDITLKESIERKPAWNGYVKSIAVHAETBS 1096  
DB 1281 RHPADNRTOJRELQOTLIDGDIIDDIISEKLEAFN--SRYEDLS-----HIAESKOI 1330  
QY 1097 SCTSLLEQWVSMRMLLIATTHADI--NHLTDSV----- 1131  
DB 1331 SLEKQL--QVLRFTDQMLQVLSGLBELQKQTLTYLTDRIADFVPOBAQIOALEISAH 1388  
QY 1132 ---VER-----OLP-----LD-- 1139  
DB 1389 LTELRLRNMRSQPLTSPESRTAGSQMDVLQKRLAEVSTKFPQKPAFQBRMDCK 1448  
QY 1140 -VLDGTALLIV---PASVNCRLRSGMKCTLLILRLQMKRELSV-- 1181  
DB 1449 RVLDGVAAELHVLVDKVDVDPDIQTHLDKCKMLKYKTSSEVGLBEVETVITGRHIVQKQOT 1508  
QY 1182 ---DEIGPL-----TEILEGV--LOAQOOLMEKTKAV--FSAPITVQL 1219  
DB 1509 DNPXGMBQULTSLKVLVNDLGAQVTEGKODERASQRLARKKKKAASISEWLSATEFLV 1568  
QY 1220 MK---EMKSDIPQYSQVLVNVCEITLOBEVIALPDQTRHSIALSGATEDKDS-METDCC 1274  
DB 1569 QKTSBGLDGLDREISMARNVLDLEKRAKDNLITIBESSAALQNLIEGSEPILEBRICV 1628  
QY 1275 ---SNSR-----HRDQ--DGVCV--LGLHLAKELCE----- 1299  
DB 1629 LNAQMSRVRTWTEDMCNTIMHNOQLEIFPDGNVAHISTWLYQVADLDELIEKFTSQOE 1688  
QY 1300 ---VEDGSMVQV-----TRRL----- 1314  
DB 1689 IVKRLVSELDANIQVENVRDQALLMNAQSSSRREIPEKLAELNRNFEKVSQHIKSAK 1748  
QY 1315 ---PIPLTLITLL-----EVSIR 1329  
DB 1749 LLIAQEPYQCLVTTTEFFGVPPSDLEKLENDIENMLKPVKHLSESDDEKMDBSAQ 1808  
QY 1330 MKONLHTEATLH-----LLITLART-----OQATVAVAGITQSIC 1367  
DB 1809 IEEVLRGEBEMLHQPMEDNKKERIRLOLLHLTRYNKAIPAIQQRKMGQLASGRSL- 1867  
QY 1368 LPLISVQSLTNGAQTSPSARKSLDAPSWGAVRLMSLMEQLKTLRYNLFLEALDFV 1427  
DB 1868 --LPTDYAVEINIKILLCMDDVELSLNVPBLNTAAYEDPSFOEDSLKNIK-----DQDLKL 1920  
QY 1428 G-----VHOER-----TLQCLNA-----VRTVOS--LACLEADHTVGF 1459

Db 1921 GEOIAVHEKOPVILASGPEAIQIRDTLTQLNAKDRINRMYSDRKSCFDR----- 1974  
 Qy 1460 ILQLSMFKEM-HFH-----LPQLMDIQVNLGYLQACTSL-IHSRKLQHYLO----- 1507  
 Db 1975 -----MEEMQFHODLNDLTQWITEABELLVDTCPAGSLDLERKARHQOELVGISS 2027  
 Qy 1508 -----NKNDGGLPSAVAVORPPSA-----ASAAPSSSQPADTEA 1545  
 Db 2028 HQPSFALLNRITGDG-----IVQKLSQADSGFLKEXKLALGNQRWDALVAEVKORQPRLKSES 2083  
 Qy 1546 SE-----OQALHTVOYGLIKITISKTALRHFTPDV----- 1576  
 Db 2084 KQVMKRYRHQJDBEITCMLTAEHAMOKRSTTELGEMIOELRDLTQEMEVHAEKLMWLNRT 2143  
 Qy 1577 COILLDQSLDLAEVNFALSFSTFTPDSEVAPSPGTLATVNVALNML 1625  
 Db 2144 LEMLDKSLSLPERD-----KISES-----LRIVNMWTKKI 2174  
 RESULT 22  
 US-08-265-967C-2  
 / Sequence 2, Application US/08265967C  
 / Patent No. 6476200  
 / GENERAL INFORMATION:  
 / APPLICANT: SABATINI, DAVID M.  
 / APPLICANT: ERDUMMENT-BROMAGE, HEDIEYE  
 / APPLICANT: LOI, MARY  
 / APPLICANT: TEMPEST, PAUL  
 / APPLICANT: SNYDER, SOLOMON H.  
 / TITLE OF INVENTION: MAMMALIAN PROTEINS THAT BIND TO FKBP12  
 / NUMBER OF SEQUENCES: 14  
 / CORRESPONDENCE ADDRESS:  
 / ADDRESSEE: BANNER & ALLEGRETTI, LTD  
 / STREET: 1001 G STREET, N.W., 11TH FLOOR  
 / CITY: WASHINGTON  
 / STATE: D.C.  
 / COUNTRY: U.S.A.  
 / ZIP: 20001-4597  
 / COMPUTER READABLE FORM:  
 / MEDIUM TYPE: Floppy disk  
 / COMPUTER: IBM PC compatible  
 / OPERATING SYSTEM: PC-DOS/MS-DOS  
 / SOFTWARE: Patentin Release #1.0, Version #1.25  
 / CURRENT APPLICATION DATA:  
 / APPLICATION NUMBER: US/08/265, 967C  
 / FILING DATE: 27-JUN-1994  
 / CLASSIFICATION: 435  
 / ATTORNEY/AGENT INFORMATION:  
 / NAME: KAGAN, SARAH A.  
 / REGISTRATION NUMBER: 32,141  
 / REFERENCE/DOCKET NUMBER: 01107,46363  
 / TELECOMMUNICATION INFORMATION:  
 / TELEPHONE: 202-508-9100  
 / TELEFAX: 202-508-9239  
 / TELEX: 197430 BBMB UT  
 / INFORMATION FOR SEQ ID NO: 2:  
 / SEQUENCE CHARACTERISTICS:  
 / LENGTH: 2470 amino acids  
 / TYPE: amino acid  
 / TOPOLOGY: linear  
 / MOLECULE TYPE: protein  
 / ORIGINAL SOURCE:  
 / ORGANISM: Saccharomyces cerevisiae  
 / US-08-265-967C-2  
 Query Match 1.4%; Score 127; DB 4; Length 2470;  
 Best Local Similarity 18.6%; Pred. No. 0.089;  
 Matches 322; Conservative 236; Mismatches 597; Indels 574; Gaps 86;  
 Qy 160 VEVADCVDKLEKELVSKYRQ-----PEELYKTEAPMTWETNGNLMTERQ-----VSR 206  
 Db 68 VNFPRILEKIFRELTSYKERKLASISLPLDLVSLR-----HELSEIEFQAIISNDINN 121

Qy 207 WFOVCLREOS-----MLEIIEIYYAYFENAPSDLLVLTQKFEQSGSQTNNHLV 258  
 Db 122 KILBLVHTKKTINRVGAVLSIDTLISFYAYTERLN-----ETSR--- 161  
 Qy 259 DETWDPFVDRIGVPSALILVEGMDI-----ESLHKCLDDBRELHQRPAQDGLICQMDCL 313  
 Db 162 -----LAGYIRGLIPSNDEVWRMLAKTKIGKLAIVPG-----GYTSPFVEFEIKSCL 208  
 Qy 314 -MLTFG-----DIPHAAPVLAMALLBHT--LNPBETSSVVRKIGT-----A 353  
 Db 209 EMLTASTEKNSFSSSKPDHAKHALLITLAEKCPVLQVYNSILINDINWRALRDPHLV 268  
 Qy 354 IQANVFOYLRLLQSLASGNDCTTS-----TACMCYGLLSFYLTSE-LH-TLGNQODI 407  
 Db 269 IRIDASITLAKCLSTLNNRDPQLTSQWVORLASCCEVG--FQVNTLECHHALLVYKEI 325  
 Qy 408 IDTFACEVLADPSLPFLFWGTPEPTSGILIDSV-----GMFPHLSPLQLLR 456  
 Db 326 L-----FLKDPFLNQVF-----DMCNCIACIENHAKAKIRKIKYQIVPBLA 367  
 Qy 457 ALVSGSTAKKVSFLDKSFLNELYKHKPHDIVSHEDGTLWRQRPKLLYPLGGQTNLR 516  
 Db 368 SFNPOLFAGKYLHQIMDN--YLEIITNAPAKKIPH-----LKDRKQLILISIG-- 413  
 Qy 517 IPQGVQVWLDRAVLVMEYSWTLFTCEIEMLHVSTADVIQHCQRKPTIDLV 576  
 Db 414 -----DIATFV-----GPDIAFY--VKQILDYI 434  
 Qy 577 HKVISTDLADIADLPITSRIYMLQRLTVISPP-----VVIVASC-----VNC 621  
 Db 435 EHDLOTRKFKFRK--KFENEIFYCIGRLAVPGVGLKLNRIIDLMRKCPISDYMGET 491  
 Qy 622 LVTLANRPA-----KWITDLRHTGFL-PRVAPVSLSGMTAEGNAGCYGN 669  
 Db 492 FOILTERIBSLGPKINDELNLVCSLTSGTPFQPSFMEIPFSFERARERMRNK---S 547  
 Qy 670 LLM-----NSEQPOGEYGVTIAPFLITLTVKQGLSGTOSGLVPCVWFVLEKMLPSYHKMR 726  
 Db 548 ILQKTESDNDNDNDIKIIOARMLKN-IKSR-----SLVEFRIVALSYI-----E 594  
 Qy 727 YNSHGVREQIGCLIELHAILNLCHETDLSHSRPSLOFLCISLAYTEAGQVITNING 786  
 Db 595 HTDPRVRKLAALVTSCE-IYKDNICIKQSLHSLNTVS-----EVLKSLA 638  
 Qy 787, IGVDTIDMMAAPRSDGAEQGGQGLIKYKLAFSVTNNYIRLKPSPNVVSPLEQALS 846  
 Db 639 ITI-----ADPLQD-----IRLE-----VLKNLPCCFD 661  
 Qy 847 OHGAGNNLIAVLAKYIYKHPA--LPLAIOLEKRLATVAPMSVYACIGNDAAAIRDA 904  
 Db 662 POLAQPDNRLPLFTAL--HDSFNIGSVAMELVGRSLSPAVYIPSI--RKLILE 713  
 Qy 905 FILRLQ--SKIEDRIKMIIEFLTVAVETOPGLIEFLNEVKDGSQSK----- 953  
 Db 714 LITKLKFTSSREKETASLCTLRSSKDVAKPIEPLNLVILPKFOQTSSTVASTAIR 773  
 Qy 954 ---EFSL---GWSCLHVLLEIDSQOQDR-----YMCPEILH 985  
 Db 774 TIGELSVGGEDMKIYKDLFPILIKTFQDQNSFKREBAALKALQOLAASSGVITDPLD 833  
 Qy 986 RAALAFALHMODRDSAMLVLRKPKEMENL--TSPLEFGLS--PSESTSEPSILET- 1039  
 Db 834 YPELL-----GILVNIILKTENSONIRQTVTLIGLIGADIDPYQKEREVTSIT 881  
 Qy 1040 -----CALIMKI--CLEIYY--VYKSLDSGLKDTLLKFSIEKRFAYMGVY 1083  
 Db 882 DISTQONAPDIDALLMGSPSNDYYTIVIHCLIKLND-----PSLSYHTYAVI 934  
 Qy 1084 KSLAVHVAETEGSSCTSLSE-----YQMLVSAWRMLLIATTHA 1122  
 Db 935 QAI-WHIFQTLGLKCVSFLDQIIPITLDVWRKTSQSLLEFFIQLCS--LITIVRH- 988



```

OY 1123 DIMHLSBVSVRROLFDVLVDGFKALLVPAVSNCU-----RLSGMKCTJLLILRLQ 1173
Db 989 -IRPHVDSIF--QAIKDPSSVAKQOTIVSVTEALSKEBGEFKVLPTLLTFLFVILEN 1045
OY 1174 WKRELGSVDEILGPYTEILEGVLAGDQOLMEKTKAKVESAFITVLQMKEMKVSIDIPQYSQ 1233
Db 1046 -----DKSSDKVLSR--RVLRLLIESPOPNIEGISH 1073
OY 1234 L-----VLNVCE---TLQEBVIALFDQTRHSIALGASATEDKDSMETDCCSRHRDQDG 1285
Db 1074 LITPKIVQMAEFTSGNLRSAI-----ITIGLAKVDVLFEMS--SRIVHSLR-- 1120
OY 1286 VCVLGHMLAKELCEYDEGDGDSLTQYTR-----LPIPLTLTLTVEVLSRKQNHFT-- 1337
Db 1121 --VLSSTTSDELSKYIMNTLSLLILQMGTSFAIFIPVINEVL-----MKGHLOHTY 1170
OY 1338 -EATLHL-----LTLARTQOGATF-----VAGAGITOSICLP--LSVYQLSNNGPAQ 1383
Db 1171 DDLTRILANVDLPFKIEANTTYDKRPEOMALAGYAK--LPINSVYLSKANSNQ 1222
OY 1384 TPSASRKSLDAPSWGYYRLSWSLMEQLKTLRYNPLBEALDFVGHQERTLOCLNAVRT 1443
Db 1228 -----RTKEDQEW-----SKRLSIQLTKESPSHALRACSNLAMYPLAKELFNT-- 1273
OY 1444 VQSLACLEADHTVFILOLSFMKEMFHLBOLMRDQVNLGVCOACTSLHSRKMLQ 1503
Db 1274 --AFACV-----WELYSQYQEDL---IGSLICIALSPLNPEHTQ 1305
OY 1504 HYL-----QNKKGDLF---SAVAQRYVRPPESAASAPSSKOPADYBESQALHTVQ 1555
Db 1310 TLNLVEMEHDDKALPIPTQSGEYAEBCAHYALMAYKEIKFTKEBNSITTESLISIN 1365
OY 1556 YGLKLILSKTLAA--LHHFTPDVCQILLDOSLDLAEYNF-----LPLASTPTPTF 1603
Db 1370 ----NQNLQTDPAIGILKHAQO-----HSLQKEMTFEKLBERMEDALAHYNREKAG 1418
OY 1604 DSEVAPSGTLLATYVVALNMNLGELDKKKEPLTQVGLASTQAEGRITLK 1652
Db 1419 DTSVSVTLGKKR-----SLHALGEW-----EQLSQLAARKKWSKLQTK 1458

RESULT 23
US-08-305-790B-3
; Sequence 3, Application US/08305790B
; Patent No. 6492106
; GENERAL INFORMATION:
; APPLICANT: SABATINI, DAVID M.
; APPLICANT: ERDUJMENT-BROMAGE, HEDIVE
; APPLICANT: LUI, MARY
; APPLICANT: TEMPEST, PAUL
; APPLICANT: SNYDER, SOLOMON H.
; TITLE OF INVENTION: MAMMALIAN PROTEINS THAT BIND TO PKBP12
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BANNER & ALLEGRETTI, LTD
; STREET: 1001 G STREET, N.W., 11TH FLOOR
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20001-4597
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/305,790B
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/265,967
; FILING DATE: 27-JUN-1994

```

Query Match	1.4%	Score 127;	DB 4;	Length 2470;
Best Local Similarity	18.6%	Pred. No. 0.089;		
Matches 322;	Conservative 236;	Mismatches 597;	Indels 574;	Gaps 86
<p>ATTORNEY/AGENT INFORMATION:  NAME: KAGAN, SARAH A.  REGISTRATION NUMBER: 32,141  REFERENCE/DOCKET NUMBER: 01107.47225  TELECOMMUNICATION INFORMATION:  TELEPHONE: 202-508-9100  TELEFAX: 202-508-9239  TELEX: 197430 BAWB UT  INFORMATION FOR SEQ ID NO: 3:  SEQUENCE CHARACTERISTICS:  LENGTH: 2470 amino acids  TYPE: amino acid  TOPOLOGY: linear  MOLECULE TYPE: protein  ORIGINAL SOURCE:  ORGANISM: Saccharomyces cerevisiae  US-08-305-790B-3</p>				
Query Match	1.4%	Score 127;	DB 4;	Length 2470;
Best Local Similarity	18.6%	Pred. No. 0.089;		
Matches 322;	Conservative 236;	Mismatches 597;	Indels 574;	Gaps 86
160	VEVADCVNLEKEVSKYKROQ-----	FEELYTKETAPWETHGNLMTERO-----	VS	206
161	VEVADCVNLEKEVSKYKROQ-----	FEELYTKETAPWETHGNLMTERO-----	VS	206
68	VNPKRILEKIRELTSDYKERKLASISFLDLVLE-----	HELSTIEPQASINDINN	121	
207	WFOVCLEBQS-----	MLBEIIFYAAYEMAPSDLLVLTKEKQGFSGRQTNRLV	258	
122	KILELVHKKNTNRVAVLSIDTLISFYATYERLN-----	ETSR-----	161	
259	DETMDDPFDRIGYRSALLVLEGMDI-----	ESLNKCALDREELHQFADGLICQDMCL	313	
162	-----LAGYRLGLIPSDVEMWRLAAKTLGLAVPG-----	GTYSDFEFEFIKSL	208	
314	MLTFG-----	DIPHHAPVLLMALLRHT--LNPEETSSVVRKIGT-----	A	353
209	EWLTSTSTKNSFSSSKRPDHAKHALLITLALENCPYLLIYQINSILINDIMPALDPRHV	268		
354	IQLNVFOYLRLTLOSLAGSANDCTTS-----	TACMCVGLSLFVLTSLE-LH-TLGNOODI	407	
269	IRIDASTILACLSLTRNRDPQLTSQWQRLATSCGY-----	FQVNTLCEIHASLLVYKEI	325	
408	IDTACEVLADPSLEPELFWGTEPTSGILILDSVC-----	GMRPHILSPILQLR	456	
326	L-----FLKDEPLNQVF-----	DOMCINCAIAYENHKAKMIREKIYQIVPLLA	367	
457	ALVSGSKTAKVYSLDKMSFYNELYKHKPHVISHEGOTLMRRQTPKLYLFGQOTNR	516		
368	SPNQLPFAKGLHQMDN--YLELTINAPAKPIPH-----	LKDDKPOLLSIG-----	413	
517	IPQGVGVMDLDRAYLVYRWESYSSWTLFCEIEMLLHVSTADVIQHCQRKPIIDL	576		
414	-----DIAYEV-----	GRDIAPV--VKQILDVI	434	
577	HKVISTDISIADCLLPITSRIYMLQLTLTVISPP-----	VDVIASC-----	VNC	621
435	EHDLOTKEKFRK--KFENEIFYCIGRLAVPGYLGKLNRIINDLMFKCPLSDYMOET	491		
622	LTVLAARPA-----	KWTDLALHTGL--PFAVHPVSSLOMSABEGMAGGYEN	669	
492	FQILTERIPSLGPKINDELLMVCSTLSGTPETIQGSEMEIPSPFRERARERRNK-----	S	547	
670	LLM--NSEQPGEGVTFIAFLRLITLTVKQGLSGTSGQGLVPCVMFVLKEMLPYHKWR	726		
548	ILQKTGESNDNDNDKIITQAFRMLAKN-IKSRF-----	SLVEFRIYALSYI-----	E	594
727	YNSHGVREQIQLLELILHALINCHETDLSSHTPSLOFLICSLAYATGAQVIVINIG	786		
595	HTDPRVKALATSGE-IYVKNONICQKSLHSINSTVS-----	EVLSKLLA	638	
787	IGVDTIDMVAQAQPSDGAEGOGQQLIKYKLAFLSVTNNVIRLAKPSRNVVSPLEQALS	846		
639	ITI-----ADPLQD-----	IRLE-----YKNIINPCPD	661	

```

QY      847 OHGAHNNILAVLAKTYHKKDPA--LPRLAIQLRLKATVAPMSVYACLGNDAAAIRDA 904
      662 POLAQNDRNLRLTLAL-----HDSFNIGSVAMLVKRLSSVNPAYTPTI-----RKILIE 713
QY      905 FLTRLO--SKIEDMRIKMIIEFLTVAVETOPGLIEPLNTEVKDGDGSK-----953
      714 LITKLFSTSSREKEBTASILCTLRSSKDVAKPYIEPLNLVLPFQDTSSTVASTLR 773
QY      954 ---EFSL---GWSCLHVLLEIDSOQDR-----YKCPPLIH 985
      774 TIGELSVGGEDMKITLKDFPLITITFODQNSFRERALKALQGLAASGVIIDPLID 833
QY      986 RAIAIALHLMODRDSAMLVLRKPKFWENT--TSPLFGTIS--PSETSEPSILET- 1039
      834 YPELL-----GLVNLKTEGONIRQVNLIGLIGALIDPRQKEREVTSTT 881
QY      1040 -----CALMKII--CLEIYY--VYKGLSDGLKDTLKKFSTIKRPAWSGVY 1083
      882 DISTEQNAPRIDIALIMQMSPSNDEYTTVVIHCLIKLKD-----PSLSYHTAVI 934
QY      1084 KSLAVVAETEGSSCTSLIE-----YOMLVSAMRMILLIATTHA 1122
      935 QAI-WEIFPTLGKCYSLFDQIITPLIDVMTCSQSLIEFYQLCS--LITIVROH- 988
QY      1123 DIMHLDTSVVRQLFLVDLGTKALLVPASVNCU-----RLGSMKCTLLILILRQ 1173
      989 -IRPHVDSIF--QAIKDFSSVAKLQITIVSVIEALSKLEGEFKLVPLTLLFLVILEN 1045
QY      1174 WKRELSGVDELIGPLLEIEGLVQADQOLMEKTKAKVSAFTYQMKEMKYSIDIPQYSQ 1233
      1046 -----DKSDKVLNR--RVLRLESFGPNLEGYSH 1073
QY      1234 L--VLNACE--TLQEVIALPDQTRHSLALGATSKDMSMETDSCSRHRDQDG 1285
      1074 LITPKIVQMAEFTSGNLRSAI-----ITICKAKVDLPEMS--SLIVSLNR-- 1120
QY      1286 VCVLGLHLAKELCEYDEGDSWLOVTR-----LPILPTLLTLEVSLRKQNLHFT- 1337
      1121 -VLSSTTSDELISKVIMNTLSLLIOMGTSFAIFPVIENEVL-----MKKHIOHTIY 1170
QY      1338 -EATLHL-----LTLARTQAGATA--VAGAGTOSICPL-LSVYQLSTNGTAQ 1383
      1171 DDLTRILNNVLPYTKILEANTTYKPAQMEADAGYAK--LPINOSVLSKAMNSQ 1227
QY      1384 TPASRSKSLDAPSWGVTRLSKSLMEQLKTRLYNPLPEALDFVGVQERTLQCLINVRT 1443
      1228 ---RTKEDWQEW-----SKRLSIQLKESPSHALACSNLASMYVPLAKELFT- 1273
QY      1444 VQSLACEADHTVGFILQLSNFMKEWHFHLPOLMRDIOVNLGYLCOACTSLHSRKMLO 1503
      1274 --ARACV-----WTELYSQOEBD--TGLCTALSSPLNPPRIHQ 1309
QY      1504 HYL-----QNKNGDGLP--SAVAQRYQRPSPASAAAPSSSKQPAATSEAOALHTVQ 1555
      1310 TLLNLVEFMERHDALPIPTQSLGEVARECHAYAKALHYKEIKIKEPENSTIESLSIN 1369
QY      1556 YGLKLISKLTAA--LRHFTPDVCQIILDOSLDLAEYNF-----LPALSTPTPTF 1603
      1370 ---NQNLQTDAAIGILKHAQO-----HNSLQKETWEKLERMEDALHAVEREKAG 1418
QY      1604 DSEVAPSGTLLATVNLVNLMLGELDKKEPLTQAVGSLTQABGTRILK 1632
      1419 DTSVSVTLGKMR-----SLHALGEM--EQLSQLARKMKVSLQTYKK 1458

```

RESULT 24  
US-09-248-796A-15028

; Sequence 15028, Application US/09248796A  
; Patent No. 6747137  
; GENERAL INFORMATION:  
; APPLICANT: Keith Weinstock et al  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN

```

; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248, 796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074, 725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096, 409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 15028
; LENGTH: 1106
; TYPE: PRT
; ORGANISM: Candida albicans
; US-09-248-796A-15028

Query Match      1.4%; Score 126.5; DB 4; Length 1106;
Best Local Similarity 18.4%; Pred. No. 0.022;
Matches 230; Conservative 182; Mismatches 396; Indels 441; Gaps 65;

QY      382 CMCYGSLSTFVL-----TSLHLTLGNQODI-IDTACEVLADPSLPE 422
      11 CVCLVQNHFEVLSEFSDYSKYTSLSFSKATNPPEVLVNMQLYLIFACSTLTATNDQK 70
QY      423 LFWGTEFTSGAGITLDSVCGMFPHLSPILQLT--RALVSGSKTAKYYSFLDKMSFYN 479
      71 ISPPNQPTH-----RKSLQMFIOHQKITSKSVKMLPPL-KLS--Q 112
QY      480 ELYKHPHVDYSHEDGTMRQTPKLLYPGQ--TNL--RIPQSTGVQVMLDRAVLYV 534
      113 PMVESVYVGLSHININILKSFLENI--PLAVEWNTNKTNPQ-----DDAFRI 161
QY      535 RMESYSSWTLPTCEIEMLHVSTADVIOHQRVAPIIDL-----YHKVISTDLSTIAC 589
      162 E-----VHIIIT-----NLTEKGLHLLITDSDIVAN 189
QY      590 LDPITRIYMLQRLTIVTSPVDVASC-----VNCITVLAARNPAKYTTDLR 638
      190 LVAIVKVNKFLS-----EPALQTFECQRLRRYFCGLENVYTGLEAKSDLEK- 239
QY      639 HNGFLPFVAPPS-----SLSQMISAQMN-----AGYGNILMNSQPO 678
      240 ---LPEFARVACENYKEMCGYDSAVTERRYNAIIRIKTERAGPSALILIEIRK 295
QY      679 GEYGVITAFRLITTLVKGQSTQSOGIVPCMFVLKEMLSYHK- RYNSHGV 733
      296 ---LQLASIKMALCSQI--KKIEIPGNLAVSVDIGLMMIQLLESNEVKIK 348
QY      734 EQLGCL-----LELHAILNLCH--ETDIHS-----SHTPSLOFL 767
      349 EIGHGASRNVIERNFEFODIISAVVORCVSQTSLSEYCLFVEIEMEKNTPEIYSDILV 408
QY      768 CICS-LAYTEAGQVINWIGIGVDITIDVMAAPRSDGEGO-----GGQGLIKTV 818
      409 CLASVITANESPO-----VAKSAILIKYFEKSLSSSESLERYTEAICSSSVYK- 459
QY      819 KLAFSVTNNVIRLKPS--NVVSPLEQALSQGHAGNNLI-----AVLAKYI 863
      460 KYLFDLSNDVONIKPESLYNRISYLTQNI-----NLIPAREALISLVNPKVT 511
QY      864 --YHK-----HPALPRLAIQLKLATV--APMS-----VYACLGNDAA 900
      512 LNYETPLPEKOSTDIHEKADSSSLMWLSNLFELTVFGSSMSQVLEWLISNNVON 571
QY      901 IRDAFLTRLOSKEI-----DMRIKMIIEFLTVAVETOPGLIEPLN-----LE 944
      572 IDKIIVEPLLOSLELKNPSFVQSGRQIIDYLAFAASDPAYIITDKLINLQRRMVPPOG 631
QY      945 VQSGDSGSK-----FSLGMS-----CLH 964
      632 NKSTAPISKDYPVYCDIRKIITQNEKQVIFSLQSLMWFLVDIVTLKNDMLIEKPLILH 691
QY      965 AVLEIIDSQODRWVCPPLHRAA--TAFHLAMOD--RDSAMLVLR-----TKRFTW 1014

```

Db 692 VSLAMD-----HYLPIVQAGATFLVHLIHALVRDDPKAQETILSRENDHITLWVY 745  
Qy 1015 ENLTSPLFGTISPPS-----ETSEPSILE-----TCALIMKIC--L 1049  
Db 746 DDLNNDKGVTPPKMDLLIRSAIKMFAVAVNIQDDMSRVSLHWATICA--VRHLACRSF 804  
Qy 1050 EIVYVVKSLDOS--LQDTLKKS--IEKRFVWSGVVSLAVH--AETEGSSCTSLLEYQ 1105  
Db 805 QVFRSLSLFDOSMLKMDLHRISNTISDEYDIQSPAQIIMTLNATLAEIDSECLIFP 864  
Qy 1106 MLVSAMRLLIATTHADIMHLTDSVVRQLFDVLDG-----TXALLVPASVNCI--- 1157  
Db 865 QLF--WCSVACLSTIH-----EQEPIEVLSTLKRISXIDDAADYVACLKRT 910  
Qy 1158 -----RLGSKCTLLILLR--QMKRELASVDEILGPTLIEGLVQADQMLEKTK 1207  
Db 911 PPOKMGKFEGLQ--QVILVGRSSSTSWPTLKFELDKL-----IVHDSLIIGMD 959  
Qy 1208 AKVPSAFITVQMKMKVSDIPQVQLVLANVCELTQEEVIALPDQTRSLALGSATEDKD 1267  
Db 960 SRVLTALL-----ANMPRF-----LHLSSENVTPETENT--GMALSQLAEDSG 1001  
Qy 1268 SMETDSCSRSHRDQDGVCVGLHLAKELCEVEDGDSMWLQVTRRLPILPTLLTLEVS 1327  
Db 1002 KSS-----LAKLIVSPDK-----RR--FRSKDFLDQS 1027  
Qy 1328 LRKONLHFTF---ATHLLLTLARTQCATVAVAGITQISICLPLLSV 1373  
Db 1028 ISCIKNAPFPEYQAGTLVLLGLLFPNKMVWKTETLGLQLI--FPLVDL 1075

## RESULT 25

US-09-085-199B-9  
; Sequence 9, Application US/09085199B

; GENERAL INFORMATION:  
; APPLICANT: Hayden, Michael R.  
; APPLICANT: Hackam, Abigail  
; APPLICANT: Huq, A.H.M. Mahbubul  
; APPLICANT: Chopra, Vikramjit Singh  
; TITLE OF INVENTION: Apoptosis Modulators That Interact with the  
; TITLE OF INVENTION: Huntington's Disease Gene  
; NUMBER OF SEQUENCES: 44  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Opedani & Larson  
; STREET: PO Box 5270  
; CITY: Frisco  
; STATE: CO  
; COUNTRY: USA  
; ZIP: 80443-5270  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Discrete, 3.50 inch, 1.44 KB storage  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: MS DOS 5.0  
; SOFTWARE: WordPerfect  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/085.199B  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Larson, Marina T.  
; REGISTRATION NUMBER: 32038  
; REFERENCE/DOCKET NUMBER: UBC-P-013US2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (970) 668-2050  
; TELEFAX: (970) 668-2052  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 756  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein

; HYPOTHETICAL: no  
; ORIGINAL SOURCE:  
; ORGANISM: mouse  
; FEATURE:  
; OTHER INFORMATION: Huntington-interacting protein  
; US-09-085-199B-9

Query Match 1.4%; Score 125.5; DB 3; Length 756;  
Best Local Similarity 20.2%; Pred. No. 0.014; Mismatches 261; Indels 181; Gaps 26;  
Matches 136; Conservative 96; Mismatch 261; Indels 181; Gaps 26;

Qy 1126 HLTDVVR-----QFLVDLDTKALLVPASVNCI-----RLGSKCTLL 1168  
Db 94 HLIERVREISGLQGLDNMKIESORAMLQKGVSEBELAQHGRAMDCELR 153  
Qy 1169 ILRQMKRELASVDEILGPTLIEGLVQADQMLEKTKAVPSAFITVQMKMKVSDI 1228  
Db 154 TELDELKQREDTEKQARSITEI--ERKAQANEQRYSKLKEK----- 193  
Qy 1229 PQVSQLVLANVCELTQEE--EVIALPDQTRSLALGSATEDKDSMET-----DDCSRRHRDQ 1262  
Db 194 --YSELVQNHADILRKNAEVTQVAVARQA--QVLEBKEKELADSFARVSDQAKQTOEQ 250  
Qy 1283 RDGVCVGLHLAKELCEVD-----EDGDSMWLQVTRRLPILPTLLTLEVSRLMKQ 1332  
Db 251 QDVLENLKHETLATSROELQVHLSNLETSAGSEAKML-----TQIAELE--KEG 297  
Qy 1333 NLHFTFATLHLLTLARTQCATVAVAGITQISICLPLSVYQUSLSTNGTAQTPSARKSL 1392  
Db 298 SLATVAQREBELALRDQLESTQIKLAGQESWC-----QVVDQQRKTL 342  
Qy 1393 DAPMPGVYRLSMLEQLKTLRLVNPFLPDLVDVGVNQHETLQCLNANRTVQSLCLSE 1452  
Db 343 LA---GIRKAEREIQEALSQI-----EEFTL-----ISCAGS 372  
Qy 1453 ADHTVGFILQSLNPMKEWHFPLPOLMRDIQVNLGCLQAC---TSLHSRKMLOHYLON 1508  
Db 373 TDHLSKVSYSVSCLEQLEKNGSQ-----YL--ACGEDISELHSTLLAHL--- 417  
Qy 1509 KNGGCLPSAVQAVQRPSPASAAAPSSKQPADTEA--SEQQLHVTYQGLKILSKTL 1566  
Db 418 -TGTPTVIGSATSRAPPEPADSLTEACROYGRETLAVLSLEEGVENADVLRNCL 476  
Qy 1567 AALRHPTDVCQILLDSLDLAENFLFALSFTPTPDSVAPSPGTLATVVALNMG 1626  
Db 477 SRVTLGEE---LLPGLDIKQEL-----GDLVDKEMAAV---SAILEATTRIE 521  
Qy 1627 ELDKKKEPTQAVGLSTQAEGRITLKSILMTFMCNCFILISQAMRYLRDPAVHRDQR 1686  
Db 522 EILSKRAGDGVGLVNNERTLIGSTSLM-----QAIKVL---VVASKDLQK 565  
Qy 1687 MKQEL---SSELSTLLSLSHYFRGAPSSPATG-----VLPSFGKSTSLSKA 1732  
Db 566 EIVSGRGSASPKFAYANSRHTEGLISASAVVGATIMVADADLVQGGKKEELMVC 625  
Qy 1733 SPESQEPILQVOA 1746  
Db 626 SREIAASTAQLVAA 639

## RESULT 26

US-08-726-214-8  
; Sequence 8, Application US/08726214  
; Patent No. 6107076  
; GENERAL INFORMATION:  
; APPLICANT: Tang, Wei-Jen  
; APPLICANT: Gilman, Alfred G.  
; TITLE OF INVENTION: SOLUBLE MAMMALIAN ADENYLYL CYCLASE  
; TITLE OF INVENTION: AND USES THEREFOR  
; NUMBER OF SEQUENCES: 31  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Arnold, White & Durkee  
; STREET: P.O. Box 4433

CITY: Houston  
STATE: Texas  
COUNTRY: United States of America  
ZIP: 77210  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/726,214  
FILING DATE: Concurrently Herewith  
CLASSIFICATION: 435  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 60/005,498  
FILING DATE: 04-OCT-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Highlander, Steven L.  
REGISTRATION NUMBER: 37,642  
REFERENCE/DOCKET NUMBER: UTSD:450  
TELEPHONE: (512) 418-3000  
TELEFAX: (512) 474-7577  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1064 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
US-08-726-214-8

Query Match 1.4%; Score 125.5; DB 3; Length 1064;  
Best Local Similarity 19.4%; Pred. No. 0.026;  
Matches 115; Conservative 74; Mismatches 166; Indels 227; Gaps 26;  
QY 966 VLELDSDQQRVYCPPLHRAAIFLALMQRDSDAMLVKTKPK----- 1012  
DB 541 VLEQNSQKQ-----WKQSDFFMLTLTYREKEMKQYRLSALPA 580  
QY 1013 -----FWENLTSPLEFGLTSPSETSEPSITELFCALIMKICLEY-----VYKGS 1058  
DB 581 FKYYACFELVFLPSFTQMLVTRPPALATTYSTIFLLFLFVCSSEHLTCVQKG- 639  
QY 1059 LDQSLKDTLKRFSEKRPAYSGVYKSLAVHAETEG-----SSCTSLLEYQMLVSAMRM 1113  
DB 640 -----PKMLHW-----LPALSVLVATRPGLRVALGTATILVFTMAV-----V 677  
QY 1114 LLITATTHADIMHLTDSVVRQLFLD-VLDGTALLLVPAVNCRLRGSMCTL----- 1166  
DB 678 SLFLPLVSSDCPLAPNVSSVAENTSWELPASLPISIPYSMHCCVGLFSLFMSF 737  
QY 1167 ----- 1166  
DB 738 ELKLLLLMLVASCSLFISHAWMLSDCLLARLYQSGLSGRPVLYKPKLMGATYFIF 797  
QY 1167 -LLILLRQ-----WKRELGSVDIELGPLTEILEGLVQ-----DQQL 1202  
DB 798 FTLLVLARQNEYCYRDLPLWTKKLRQERE-----ETETMENTLPAHVAPOULIGNRRNEDL 853  
QY 1203 MEKTKAKFASAFITVLQWKEKMSVDIPQYSQVLN-----VCETLQEBVIALFPD----- 1252  
DB 854 YHGVCECVLPFAIIPPEKFE-----YSSNINHEBIEGLRLNELLADFDLSSKPK 906  
QY 1253 -----TRHSLAIG-SATEDKQSMETDCCSR-SRRHDDRGVCV-LGLHLAKELC 1298  
DB 907 FSGVEKIKTIGSTYMAATGLNATPGQDTQ--DAERSCSHL-----GTMVEPAVALGSKLG 960  
QY 1299 EVDDEGSMQVTRRL--PILPTLL-----TILEVSLRMQ-----NLHTEA 1339  
DB 961 VINHGSFNNFRLRVLGHGVPVAGVIGAKQFOYDIMGNTVAVASRMESTGLKICVTEE 1020  
QY 1340 TLHLLTLART--QCGATAVAGAGITQSICLPLTSVYQLSTNGR-AQTPSAS 1388

DB 1021 TARALQSLGTCYRGVIRKYGKG-----QLCTYFLATDLTRTSSPAS 1064  
RESULT 27  
US-09-085-199B-4  
Sequence 4, Application US/09085199B  
Patent No. 6235879  
GENERAL INFORMATION:  
APPLICANT: Hayden, Michael R.  
APPLICANT: Hackam, Abigail  
APPLICANT: Huq, A.H.M. Mahbubul  
APPLICANT: Chopra, Vikramjit Singh  
APPLICANT: Kalchman, Michael  
TITLE OF INVENTION: Apoptosis Modulators That Interact with the  
NUMBER OF INVENTIONS: 44  
TITLE OF INVENTION: Huntington's Disease Gene  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Opedahl & Larson  
STREET: PO Box 5270  
CITY: Frisco  
STATE: CO  
COUNTRY: USA  
ZIP: 80443-5270  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Kb storage  
COMPUTER: IBM compatible  
OPERATING SYSTEM: MS DOS 5.0  
SOFTWARE: WordPerfect  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/085,199B  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Larson, Marina T.  
REGISTRATION NUMBER: 32038  
REFERENCE/DOCKET NUMBER: UBC.P-013052  
TELEPHONE: (970) 668-2050  
TELEFAX: (970) 668-2052  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 914  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHEICAL: no  
ORIGINAL SOURCE:  
ORGANISM: human  
FEATURE: Huntington-interacting protein  
US-09-085-199B-4  
Query Match 1.4%; Score 125; DB 3; Length 914;  
Best Local Similarity 21.6%; Pred. No. 0.022;  
Matches 134; Conservative 76; Mismatches 244; Indels 166; Gaps 27;  
QY 1164 CTLLILRLQWRELGSVDIELGPLTEILEGLVQADQQLMEKTKAVFSAFITVLQWKEM 1223  
DB 307 CEFLRAELDELRLQRQREDTQKQRSLEI-ERRAQNEGRYSLKKEK----- 351  
QY 1224 KVSIDIQYQVLANVETIQE--EYIALPDQTRHSLALGASATEDKQSMETDCCSRHDD 1281  
DB 352 -----YSELVQNHADLLRKNAEYTKQVSMARQA-----QVDLERKEKLEED-SIERISD 399  
QY 1282 QRDGVCVLGLHLAKELCEVEDGDSWLQVTRRLPILPTLLTTLVSLRMKQHLHTEATL 1341  
DB 400 Q-----GQRTQGBLEVLLESILKQELGTSQR--ELQVLOGSLFYSAGSEANMAAFAPAL 450  
QY 1342 HLHLLTLARTQAGATVAGAGITQSICLPLLSYQLSTNGTACTPASRSKSLDAPSPGVY 1401  
DB 451 E-----KERDSIVSGAARE-----BELSALREKLEQDT----- 478  
QY 1402 RLMSLME-----QLKTLRYNPLPALDPVGVHQ--ERTLQ--CLNAATVQVSLACLEAD 1454

```

Db      479 QLKLAESTEEMCOLAKDQRKML-----VSGRKAAGVIGDALNQLPEEPILSCAGSAD 532
Qy      1455 HTVGITILQSNFM-----KEMPHLPQMRDIOVNLGYLQAC-----TSLHNRKMLQHYL 1506
Db      533 HLLSTVTSISSCIEGLEKSWSQYL-----ACPEDISGLHSTITLALH- 575
Qy      1507 QNKINQDGLPSAAGVQRPSPASASAPSSKOPADTEA-----SEOOALHTVOYGLK- 1560
Db      576 ---TSDALAHGATTCILRAPPEPBDLTCEKQYGETTAYLASLEBESGLEADSTAKRN 632
Qy      1561 ILSKTIALRHPTDVCQIILLDOQSLDLAEYNLPAISFTTPTPDESVAPSGTLLATVNV 1620
Db      633 CLSKIKA-----IGEBLLPRGLDIQEBEL-----GDLVDKEMAT-----SAAIET 673
Qy      1621 ALNMGELDKKEPELTQAVGSTQAEGRITLKSLLMTFMENCFYLLISQAMAYLADPAVH 1680
Db      674 CTARLEEMLSKSRAGDTGVKLENNERILNCTSLM-----QALQVL---IYA 717
Qy      1681 PRDKORMKQE-----LSSELSTLSSLSRYFRGAP--SSPATGVLPSPQGS 1726
Db      718 SKDLQREIVESGRGTASPEKPEFYAKNSRWTEGLISASKAVGATVMDAADLVVGRGKF 777
Qy      1727 TSLSKASPESGEPILLQVQA 1746
Db      778 EELMVCSEHIAASTAQLVAA 797

RESULT 28
US-09-228-246-2
; Sequence 2, Application US/09228246
; Patent No. 6245510
; GENERAL INFORMATION:
; APPLICANT: Staerkawicz, B. S. et al.
; TITLE OF INVENTION: PRP Protein and Nucleic Acid Sequences: Compositions
; FILE REFERENCE: 51700
; CURRENT APPLICATION NUMBER: US/09/228,246
; CURRENT FILING DATE: 1999-01-11
; EARLIER APPLICATION NUMBER: 08/680,327
; EARLIER FILING DATE: 1996-07-11
; EARLIER APPLICATION NUMBER: 08/310,912
; EARLIER FILING DATE: 1994-09-22
; EARLIER APPLICATION NUMBER: 08/227,360
; EARLIER FILING DATE: 1994-04-13
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: Patent In Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1824
; TYPE: PRP
; ORGANISM: Lycopersicon esculentum
US-09-228-246-2

Query Match 1.4%; Score 124.5; DB 3; Length 1824;
Beet Local Similarity 19.0%; Pred. No. 0.09;
Matches 305; Conservative 209; Mismatches 552; Indels 541; Gaps 81;

```

```

Db      441 LLLKFKFKVAILRQKCPISASSTPDHMDILNLPNNFELIDSSYSSMLKASCPSSSHR 500
Qy      253 TNRHLVDETMD---PEYDRIGYSALILVSGMD--ISLHKCAL-----DDRR----- 295
Db      501 PNRDAESPNTSEFLCPNPDVVSFYSSSRIPRPMDBILKRFHEYILLNLRKXETMLTFTLI 560
Qy      296 --ELHQFQPDGLIQODMCMILTF---GDIPHAVALMALNRLHTLNPBESSVY----- 346
Db      561 ADEVKRF-YDGLL-----LWVTLIEPPVPHTECKRQNDLSMRHEAVALIESAVCLHY 613
Qy      347 -RKIGTATQLN-VFQVLTRELQSLASGANDCTTSTACVCVGLSFVLTSTHELN--TLG 402
Db      614 EDNMNNNRREINQVLFPLTVFWLKSNG-----LMDLKKKSTLG 655
Qy      403 NO-QDIIPTAGE-----VLAD-----PSLPELPWGTPTSGGILLDSVC-----G 442
Db      656 NOVLDLIESAHBELILNLSIIMDLRKLYRLDLMHABVYAKRLAIFSGSCYEYFVNG 715
Qy      443 MPEHLASPLL-QLRALVSGSTAKV-----YSPLDKMSFYNELYHKH---PHDV 489
Db      716 SSTEMRPLSDPFLQELIESVKVEFRNVCLOVLDISPFLTDBEGVLNFKKQAVPND 775
Qy      490 ISHEDGTLM-RROTEKLLYP---LEGQTNLRIPQG-----TVGQ----- 524
Db      776 AVSSDGSLEDASTTEKMLPSDFLEISVEIKERARKYDQVLDATHCETSKTDGKSFIN 835
Qy      525 VMLDDRVLVWREVSYSWTLFTCEI--EMLIHVS-TADVIGHCQRYKPIIDLVHAKYIS 581
Db      836 IMLTQODKLPDYDASVSAYLLNQISVVDKLIHISGLSLVDVIOYQNMHIELDLERVOD 895
Qy      582 TDLSTADCLLPITSNI---VWL-----LQRLTVISPPVDVYASVNCCLTVLAANPAKV 633
Db      896 KNYI---CPFSVKGYIPAWYTLTYLSDVKQLKFEABEVKII---CLKV----- 938
Qy      634 WTLDRHTGFLPFAVPVSLQMSISAEQMGNGYGNLMMN--SEOPQGEYGTIAFLRLI 691
Db      939 -----PDSSTSYSPKTNL---GYLNCILGLKEILLERKDLIDILKQI 980
Qy      692 TTLVKQGL-----GSTQSQGLVPCVM-----FVLKEMLPSTYHKRNVSH 730
Db      981 ESVKEGILLCLRSFIDHFESEYDEHDEACGLIARVSVMAYKAEVYDSCIAVSHPLMYKVL 1040
Qy      731 GYREIQGLILELHAILNLCHET-----DLHSHTPSLQFICISLAVTEAGQYVI 782
Db      1041 WISE-----VLENIKLVNKVGEETCERNIEVYAEV-AKTTTVAPSPSALTQANBEM 1094
Qy      783 NIMGIGVDTI-DMWMAQPRSD-----GAEGQSQGLIKTVKLAFSVTNNVIRLRKPPSN 836
Db      1095 EGFQDTIBELKDKLIGSPBELDIVISYGMPLGK----- 1128
Qy      837 VVSPLEQALSOHGAHGNLIVLAKYIYHKDPAIRLAIOLKRLATVAPMSVYA-CLG 895
Db      1129 -----TTIAKKIY--NDPEV-----TSRFDVAQCVV 1153
Qy      896 NDAALIRDAFLTRLOSKIE-----DMRITVMILEPITVAVENQPGILFL 941
Db      1154 TQLYSWRELTLITLNDVLEPSDRNEKEDJADELRFLTLTRFLI-----LIDVW 1205
Qy      942 NLEVWDG-----SDSKFESLGMMSCILHAYVELIDSOQOODRYWCPPLHRAAIAFLHLM 996
Db      1206 DYKWDNLCMCFPSDVSNNSRIILITRLNDVAEYVCESDP-----HHLR 1249
Qy      997 QDRDSAMLVLRKPKFENITLSPFLGTLSPSETSEP--SILETC-----AL 1042
Db      1250 LFRDDDS-----WTLQKEVFQGESCPBELEDVGFRISSCRGLPLSVLVAVGL 1299
Qy      1043 IMKIICLEIYYVK-----GSLDOSLQDTIKKES-----IEKRFAYSGVYKSLA 1087
Db      1300 KQKKTTLDSMKVNEOSLSSQRIGSLESI--SIFGSYKMLPHYLKPCFLPFGGFLQGD 1357
Qy      1088 VHAETEGSSCTSLLEYQMLVASAMRMLLIATTHADIMHLLTDSVVRQLFLDVLDTGTAL 1147

```

Db 1358 IHVSK-----MTKLM-----VAEGFVQANNKGEQEDTAQGLDLDLIGRNV 1398  
QY 1148 LIPVSVN-----CLRLSGMKCTLLI-----LIRQWRELGVS-D 1182  
Db 1399 NAMEKRPNTKVKTCRIHDLHKFCMEKAKQEDPLQINSGEVFPFERLEEYPLFVHSYOD 1458  
QY 1183 EI-----LGPLTEILEGVLOADQOIMETKAKVFSAPITVLQMKEMKVSIDIPOYSQVL 1236  
Db 1459 EIDLWRPSPRSNVSILFNADIPDNLMPRDISFIFESF-----KLVKYLDLESF----- 1507  
QY 1237 NVCEFLQEEV-----IALFDQTRHSIALGASATEDKDSMTD-----DCS----- 1275  
Db 1508 NIGGFPEETIQYLIQMKFPAAQTDANSIPSSIAKLENETTFVVRGLGEMILPCSLMKV 1567  
QY 1276 RSRHDPDRGVCVGLGLHLAKELCEVEDDGSWLOVTRRLPILPTLLT 1322  
Db 1568 KLRHIVANDRVS-FGLH-----ENMDV-LTGNSQLPNETST 1603

RESULT 29  
US-09-085-199B-5  
; Sequence 5, Application US/09085199B  
; Patent No. 6235879  
; GENERAL INFORMATION:  
; APPLICANT: Hayden, Michael R.  
; APPLICANT: Hackam, Abigail  
; APPLICANT: Chopra, Vikramjit Singh  
; APPLICANT: Kalchman, Michael  
; TITLE OF INVENTION: Apoptosis Modulators That Interact with the  
; TITLE OF INVENTION: Huntington's Disease Gene  
; NUMBER OF SEQUENCES: 44  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Opedahl & Larson  
; STREET: PO Box 5270  
; CITY: Folsco  
; STATE: CO  
; COUNTRY: USA  
; ZIP: 80443-5270  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 KB storage  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: MS DOS 5.0  
; SOFTWARE: Wordperfect  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/085.199B  
; FILING DATE:  
; CLASSIFICATION:  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Larson, Marina T.  
; REGISTRATION NUMBER: 32038  
; REFERENCE/DOCKET NUMBER: UBC-P-013US2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (970) 668-2050  
; TELEFAX: (970) 668-2052  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1090  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHEICAL: no  
; ORIGINAL SOURCE:  
; ORGANISM: human  
; FEATURE:  
; OTHER INFORMATION: Huntington-interacting protein  
US-09-085-199B-5

Query Match 1.4%; Score 123.5; DB 3; Length 1090;  
Best Local Similarity 22.0%; Pred. No. 0.043;  
Matches 99; Conservative 61; Mismatches 164; Indels 125; Gaps 21;  
QY 1164 CTTLLILRLQKRELGSVDELIGPLTEILEGVLOADQOIMETKAKVFSAPITVLQMKEM 1223

Db 483 CEPLRAELDELRRQREDTEKQKRLSEI-EKKAQANQRYKLEK----- 527  
QY 1224 KVSIDIPOYSQVLAVNCEITLQ-EVIALFDQTRHSIALGATEDKDSMTEDDCRSRRHD 1281  
Db 528 -----YSELVQNHADILRRKNAEVTQVSWARQA-----QVLDREKKELED-SLERISD 575  
QY 1282 QRDGVCVGLHHLAKELCEVEDDGSWLOVTRRLPILPTLLTLEVSIMKONLHFTETL 1341  
Db 576 Q-----GQRKTEQLEVLSEKQELATSOR-ELOYLGSLTSSAQSENNMAAFEL 626  
QY 1342 HLLLTARTQGAFAVAGAGITGOSICUPLLSVYOLSTNGTASAKRSIDAPSPGVY 1401  
Db 627 E-----KERDELVSQAHR-----EELSARKELQDT----- 654  
QY 1402 RLMSLWE-----QLKLRNLFPLDEALDFGVHO-ERTLO-CLNAVTVQSLACLEAD 1454  
Db 655 QKLASTEESMCQIAKQORRKL-----VGSRRQAQVIODALNQLSEEPPLISCAGSAD 708  
QY 1455 HTVGFILOLSNFM-----KEMHFLPOLMRDIOVNLGYLQAC-----TSLSRRKMLQHYL 1506  
Db 709 HLLSTVISISCIQLKRSQYL-----ACPEDISGLHSTILLHL- 751  
QY 1507 QNKXGDLPSAVAGVORPPSASAPSSSKOPADTEA-----SEQOALHTVQYGLK- 1560  
Db 752 ---TSDAIAAGATCTCLAPPEPADSLTEACKQYGRETLAVLASLEGSLENNADSTAMRN 808  
QY 1561 ILSTTLAALHRTFDVQOIIIDGSLDLAE 1589  
Db 809 CLSKIKR-----IGEBLLPRGLDITQ 829

RESULT 30  
US-08-680-327-3  
; Sequence 3, Application US/08680327  
; Patent No. 5853521  
; GENERAL INFORMATION:  
; APPLICANT: Staskawicz, Brian S., Oldroyd, Giles Edward,  
; APPLICANT: Salmeron, John M., Rommens, Cais  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR PLANT  
; TITLE OF INVENTION: PATHOGEN RESISTANCE  
; NUMBER OF SEQUENCES: 15  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Klarkquist Sparkman Campbell Leigh &  
; ADDRESS: Winston  
; STREET: One World Trade Center  
; STREET: 121 S.W. Salmon Street  
; STREET: Suite 1600  
; CITY: Portland  
; STATE: Oregon  
; COUNTRY: United States of America  
; ZIP: 97204  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Disk, 3-1/2 inch  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: MS DOS  
; SOFTWARE: Wordperfect 5.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/680,327  
; FILING DATE: July 11, 1996  
; CLASSIFICATION: 800  
; PRIOR APPLICATION NUMBER: 08/310,912  
; FILING DATE: September 22, 1994  
; CLASSIFICATION: 800  
; APPLICATION NUMBER: 08/227,360  
; FILING DATE: April 13, 1994  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Dow, Alan E.  
; REGISTRATION NUMBER: 35,123  
; REFERENCE/DOCKET NUMBER: 5151-45038  
; TELECOMMUNICATION INFORMATION:

TELEPHONE: (503) 226-7391  
 TELEFAX: (503) 228-9446  
 INFORMATION FOR SEQ ID NO: 3:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1824 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 US-08-680-327-3

Query Match 1.4%; Score 123.5; DB 2; Length 1824;  
 Best Local Similarity 18.6%; Pred. No. 0.11;  
 Matches 300; Conservative 206; Mismatches 558; Indels 545; Gaps 77;

40 LNKHRRRLLEGSLVYKPPSP-----SSAEKVYKXK-----DVASPLK-ELGL 80  
 214 MDDESKLERIS--RPPSPGLRQYVLAALGLNSRSKTYMAKYMADPFSALODDLRL 271  
 81 RISKFLGLDEEOSVOLLOCYLOEDYGRTRDSVKTVALODEROSQALILKIAD----- 131  
 272 RC-----DNRIWLGRLGSLYLCRPLRD-IESYPVSHROLISQLMMEDIAISANAI 322  
 132 YYEERTILNCVHLHLYFPQDERHPRYE-----YADCVDKLEKEIV- 174  
 323 YSYDEDMKTESEIDHEFLHQLMKFNYYVVEVDLIRLQNIQGTIIIVPMKDLIDYWEELMF 382  
 175 --SKYRQOFEE-----LYKTEAPTWETHGNL-----MTEROYSRMFWOGLAREOS 216  
 383 FRSFMDAPDQKEQETRIYVILNVIQSAVSOAMVCSLSLCHDNLONDLAR-BINCLHFO- 440  
 217 MLLEIIFLYAVAFEMAPS-----DLVLVTKMFKEQ-----GFGSRQ 252  
 441 LLKRFKIKVALRQMCPSIASSTPDHIMIDLNLFPNPFALIDSYSSMLKASCSSSHR 500  
 253 TNRHLVDSTMD---PFVDRIGYFSALLVBSGMDIESLHKCALDRRLHOPADGLICQ 308  
 501 ENRDAESNPTSFELCGPNTDVYSFYSSSRIPKMD-----ELIKFHEHIIIVNLARK 551  
 309 DMDCLMLTFEGD-----IPHNAPVLAMALIRHLINPESTSV-- 346  
 552 DETNLFTFIADBYKKFYEGGLIMTYLLIEPPVPHTECKKONDLSKRHEAVALEASAVCL 611  
 347 --RKIGGTALQNL-VFOYLTRLLQSLASGNDCTTSTACMCVYGLSFVLTSLELH--T 400  
 612 HYEDMNNNSREINQVLFVTFPMLIKSEGN-----LMDLKHST 653  
 401 LGNQ-QDIIIDTACE-----VLAD-----PSLEFLWGEPPISGLGIILDSVC----- 441  
 654 LGNQVLADIESAHEELILRSILMDLRLKQLYRLDLMHAEVTAKRLAISGSCYEYEM 713  
 442 -GMPHLLSPLL-QLLRALVSGSKTAKKY-----YSFLDKMSFYNELYKHK---PH 487  
 714 NGSSTEKRRPLSDPLQETIESVYKVEFRVNCLOVLDISPSSLTDSGLVNFLLKONAKYPN 773  
 488 DIVSHEDSTLW-RROTPKLLYP---LGGOTNLRIPOG-----TVGO-- 524  
 774 DDAYSDDSLBASSTKMGLEPSDPLREHSEVEIKKAKLYQVLDATHCHETSXTDGSF 833  
 525 --VMLDRAVILRWMEYSYSWTLFTCEI--EMLLHVVS-TADVIOHCORVKEPIIDLVKV 579  
 834 INIMLTQODKLPDYDAGSVSYLLNQISVVKDLHLIGSLAVDIQVYRMHIEITLTLARV 893  
 580 ISTDLSINDCLLPISRI---YML-----LQRLTVISPPVDVVASCNCLTVLAARPA 631  
 894 QDKNTI---CFPSVGVYIPAWYTYLSDVKQLKFEVEVKII---CLKV----- 938  
 632 KVTMTLRHTGFLPFAHVVSSLSQWISAGMNAAGVGNLLMN--SEPOGEGVYTIARL 689  
 939 -----PDSSSYSPKTNGL---GYLNCFLGKLELLRSKLDLIDLKH 978  
 690 LITLVKCOL-----GSTOSQGLVPCVM-----FVLKEMLPYHKMRYN 728  
 979 QIESVKEGILCLRSFIDHFSESYDEHDERACGLIARVSVMAVKAEEYVIDSCIAVSHPLWYK 1038

729 SHGREQIGCLLELILHAILNLCHET-----DLHSHRPSLOFLCISLAATEACOT 780  
 1039 VLMISE-----VLENIKLVNVVGETCERNIEVYHEV-AKTTTYVAPPSATYGRANE 1092  
 781 VINIMIGIVDTI-DMVMAAOPRSD-----GAEGGQGGOLLIKTYKLAFSVTNVNIIRLKP 834  
 1093 EMEGFQDTIDELKDKLGGSPELDVIISVGMPLGAK----- 1128  
 835 SNVVSPLQALSOQHAGHNNNLIATKAYIYKHDPALPRLAIQLKRLATVAPMSVYA-C 893  
 1129 -----TTLAKKIY--NDPEV-----TSRPDVHACQ 1151  
 894 LGNDAAAIRDAFLTRLOSRIE-----DMRKIMILEFLTVAVETOPGLIEL 939  
 1152 VVTQIYSWRRELLITLINDVLEPSDRNEKEDBIADLRFLITRKFLI-----LIDD 1203  
 940 FLNLEVDKG-----SDGSKFSLGMSCLHAIVLELIDSQODRWYCPPLHRAATAFLHA 994  
 1204 VMDYKVMNDLQMCPSDVSNRSRIILITRLINDVAEYVKCESDP-----HH 1247  
 995 LMQDRDSAMLVATKPKFWMENLSPLEGTLSPPSETSEP--SILETC----- 1040  
 1248 LRLFRDDES-----WTLQKEVFOGSSCPLELEDVGFELSKSCRGPLSVLVLAG 1297  
 1041 ALIMKIICLEIYYVK-----GSLDQSLKDTLKRS-----IEKRFAYMSGYYS 1085  
 1298 VLKQKKTLDMSKVVESLSSORIGSLSESI--SIGFSYKNLPHVYLKPCFLYRGGFIOG 1355  
 1086 LAVHVAETEGSSCTSLLEYQMLVSAWRMLLIATTHADIMHLTDSVVRQLFLVDLGTK 1145  
 1356 KDHIYSK-----MTKLM-----VAEGFVQANNEKQEDPAQFLDDLIGRN 1396  
 1146 ALLVPAASN-----CLRSGMKCTILLI-----LQMKRELGSV 1181  
 1397 VVMAMEKRPNKVKTCRIHDLHAKFCMEKAOEFLQINSGEVFPERLEBYRLFVHSY 1456  
 1182 -DEI-----LGPTLEILEGLVADQOQMEKTKAKVFSAFTVLQMKEMKYSDIPOYSOL 1234  
 1457 QDEIDMPPRSNSNRSLFPMNIDPDNLMPDIDISPIFSF-----KLVTLDLESF-- 1507  
 1235 VLVNCELTQEEV-----TALFDQTRHSIALGASATEDKDSMETD-----DCS-- 1275  
 1508 --NIGSTPTEIOYLQMKYPAQTDANSIPSSIAKLENLETFVVRGLGEMILLPCSLIK 1565  
 1276 --RSRHRQDRGVCYLGLAKELCEVDEDEDSMLQVTRRLPIITLLT 1322  
 1566 MVKLRHIVNDRVS-FGLH-----ENMDV-LTQNSQLPNIETEST 1603

RESULT 31  
 US-09-542-331-2  
 ; Sequence 2, Application US/09542331  
 ; Patent No. 6261761  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Zhong, Yi  
 ; APPLICANT: Guo, Hui-Fu  
 ; TITLE OF INVENTION: NF1 Protein and Its Role in Activation  
 ; TITLE OF INVENTION: of Adenylyl Cyclase by PACAP38-like Neuropeptides  
 ; FILE REFERENCE: 1314.1047003  
 ; CURRENT APPLICATION NUMBER: US/09/542,331  
 ; CURRENT FILING DATE: 2000-04-04  
 ; EARLIER APPLICATION NUMBER: US 09/046,745  
 ; EARLIER FILING DATE: 1998-03-24  
 ; EARLIER APPLICATION NUMBER: US 60/041,469  
 ; EARLIER FILING DATE: 1997-03-24  
 ; NUMBER OF SEQ ID NOS: 3  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 2  
 ; LENGTH: 2818  
 ; TYPE: PRT  
 ; ORGANISM: Human  
 ; US-09-542-331-2

Query Match 1.4#: Score 123.5; DB 3; Length 2818;  
 Best Local Similarity 18.5%; Pred. No. 0.26;  
 Matches 380; Conservative 256; Mismatches 698; Indels 717; Gaps 101;

QY 63 EKVANDVNSPLKELGLRISKFLGDEBQSVLLQCYQEDYKGT-----DSVTVLQ 117  
 DB 1020 EVMARDDJDSFCQEMKFR-----NMKVE-----YLDWMWGSNQAADDVACLR 1066

QY 118 DERGQALILKLAIVYEEERTCLACVLLHLYTQDERHRYRYEYVADCDKLE--KEVLS 175  
 DB 1067 DLDQAS-----MEAVSLIAGL-----PLQPEEGDGVLEMAKSQLFL 1104

QY 176 KYRQOEELY-----KTEAPTWETHGNLITEROVSFMEVQCLRE-----QSM 217  
 DB 1105 KYFTLFNNLINDCEVEDESAQGTGR--KRGSRLAS--LRICTYLANSNLINANVDSG 1160

QY 218 LLELIEL--YVAFEMASDLYITKMFKEGFGSSRQTNRLVDETMDPVDRIQVFSALI 276  
 DB 1161 LMSISGLGHKDIQTRATEFMEVLTKL--QOG-----TEFDLAEIV--LADRFERLVELV 1212

QY 277 LVEG-----MDISLHKCA-----LDRREHQ-----PAQDGLLQCDMD 311  
 DB 1213 IMMGOCELPFAMALANVPCSQMDLAVLYTLFDSRHLYQLMNNFSKEVELADSQ 1272

QY 312 CL-----MLTEGDIPIHAFV--LAMALIRHTLNPEETSSVVRKIGTAI----- 354  
 DB 1273 TFRGNSLASKIMTFCKFYAGATYLOKLDPLIRIYITSSDWQHSFEVDPRLEPSSSL 1332

QY 355 ---QLVFOYLTRLQSLASGNDCTTS--TACMCYGLSLFVLSBELHTIG----- 402  
 DB 1333 EENQNRULQHTERFPAHAISSSEFPQARSCHLYQVQSPONSIGAAGSAMFLRF 1392

QY 403 -NOODIIDPACVYLDPSLPELFWGTEPTSGIITLDSVCGMFPHLSPLQLLRA----- 457  
 DB 1393 IMPAIVSPYEAGLLDKKPPRIERGLKLSK--ILQSIAN--HVLFTKEHMRPRDF 1446

QY 458 LVSGSTAKKVVSPFDKMSFYNELYKRP--HDVISH-----EDGR-----LMRQTPK 504  
 DB 1447 VASNPDAAR-----FPIIDIASDCPTSDAVNHSLSFISDGVNLAHRLNNQD-- 1495

QY 505 LIYPLGGQTNLRIPQTVGOVMLDRAVILVRW-----EYSSWTLFTCEIB--M 552  
 DB 1496 ---KIGQVLSNRDKHAGVRPRFDKQATLAVLGRPHKPVADHTMSSNLTSKFEFPM 1552

QY 553 LHVSTADVLQHCGRV-----KPIIDLVAKVISTDLSADCLPITSRIYML 600  
 DB 1553 TRHQVHEKEEFKALKTLSIFYOAGTSKAGNPFYVVARFRFTQOINGDL-----LYHV 1606

QY 601 LQRLTTVISPPVDVIVSCNCLTVLAANPAKWTDLEHTG-----PLP--FVAHNV 650  
 DB 1607 LITLKPYYAKPYEIV-----VDLHTGPSNRKFTDLSKMFVVP-- 1646

QY 651 SLSLGMIAEGMNGYGNL-----LMSQPOQEGYVGTIAFLRLITTL-----VK 696  
 DB 1647 ---GPAVNSAVVYIYCNNSWVREY--TKYHERLITLGLKSKRLVFDICP 1691

QY 697 GOLGS--TOSQGLVPCVWFVLKEMLPYH--KVRYNSHGVRBOIG----- 737  
 DB 1692 GKLAENIHEGOQKRPATLALBEDLVKFNHMLKLAHQDTKYSIKVGSITAVQVTSABRTKY 1751

QY 738 ---CLILR-----LHAIINL----- 750  
 DB 1752 LGQSVFLNDIYVASEIEICLVADENOPTLITANOGTFLTFMHQCECAIVOSIHIRTRME 1811

QY 751 CHETDLHSHTPSLOFLCISLAYTEAGQVYINIMGIVDTID-----MMAAQPS 802  
 DB 1812 LSGQDSIPQHT-----KIRKQVPGTILNIALNLSSDPSLSAAYNLCLTCT 1862

QY 803 DGABGQGGQLLITVLAFSVTNNV-----IRLPPSNVVSPLQALSGHAGANN 854  
 DB 1863 FNLIKIEGQ---LLETSGELCI PANNTLFTIVISKTLANEPHLTLEFLBEGISFSSSIE 1919

QY 855 LIAVLAKYIYHKDPRALPRLALQILKRLATVAPMSYVACLGNDAA-----AIRDAFLT- 907

DB 1920 LKHLCLLEYM-----TPWLSNL-----VAFCKANDAKQORATLIDDKITM 1960

QY 908 ---RQSK-----IEDMRKVMLEFLTVAVETQPLI--EFLNLEVQGS 949  
 DB 1961 TINEKQWYPSIQAKWISLGOITDL-LDVVIDSFITKSATGIGSIKAEVADPAVALAS 2019

QY 950 DGSKEPSIGMWSCHAAVELLIDSGQQRWYC---PRLHHAALAFHAAQDRDRA--M 1004  
 DB 2020 GNVKLVS--KVIGRCKIITDKT-----CLSTPTLEO-----HLMDDIATILARYM 2064

QY 1005 LVATKPKFMEVNTSP--LFGTLSPSETSEPSI--LETICALIMKI-----CLEIYYVVKG 1057  
 DB 2065 LMLSPNNSLDVAALPYLFFVVTFLVATGPBLSRASHGVLVINIHSLSCTQSLHF--SE 2122

QY 1058 SLDOSLQDKKPSIEGRFAPWS--GVYKSLANVALEGGSCSLSEYOMLVSAMRLI 1116  
 DB 2123 ETKQVLAISLTERSLRPFYLLPFGISKVSAV----- 2156

QY 1117 IATTHADIMELTDSVVRQFLDVLQD--TKALLVPAVNCIR--LGSMTLLILLRQW 1174  
 DB 2157 FRSSYRDRSFPSSYERETPALTSLTETVTEALLEIMEA--CMRDIPTCK-----WLDQW 2208

QY 1175 -----KR--ELGSVDEILCEPTEILEGVLOA-----DQ 1201  
 DB 2209 TELAQRAFQYNSLOPRALVFGCISKVSHQIQIIRILSKALESCIKGPDYNSQV 2268

QY 1202 LMEKYAKVSAFITYLQMKEMKVS DIPQVSOVLVNCETLQ--EEV-----IALPDOR 1254  
 DB 2269 LIBAT-----VIALTKLOPLNKDPSLHKALFWAVAVLQDDEVNLYSAGTALLEQYL 2321

QY 1255 HSLALSGATEDKMSMETDCSSR--HDDQDQVCVLGLHAKELCEVEDGDSMLQYT 1311  
 DB 2322 HTLDSLRIFNDSPSEEVFAIRNPLEWHCKQD--HF----- 2356

QY 1312 RRLPILPTLLTLEVSLRKNQNLHF-----TEATHLILLTLARTQ 1351  
 DB 2357 -----VGLNNSNFNPAVGHLLKGRHSPRAVARTIRLHLTLVNRH 2402

QY 1352 QGATVAGAGITQISICLPILSYOULSTNGTAQTPASRSKSLDAPSWPGYRLSMSLEQL 1411  
 DB 2403 RNDCKREVN--TOSVAY--LAALLTVSEVRSCLKHKSL-----L 2441

QY 1412 LKTLRYNFPREALDFGVH---QERTLOCLNAVTVQ--SLACLEADHTVGTILOSNF 1466  
 DB 2442 LNDISMEVNP--MDYTPRHGDPSTRILAKETQPMSSPGSGYLDAATYPTVG---QTS-- 2494

QY 1467 MKEMENHLPQLMRDIQVNLGYLQACT--SLHSRMLQHYLQNKXGDLPSAVAQVQRP 1525  
 DB 2495 -----PRAKSMSLMDGQPSQANTYKLLGTRKSPHLISDTPK-----APKQEM 2538

QY 1526 PSAASAPSSSQPADDEASQOALHTVOYGLKILSKITLALBNHFPVDVQILLDQSL 1585  
 DB 2539 ESGITPPRMRNVAVETDIEMETORLSSQOHPRKYS----- 2576

QY 1586 DLAEYNLPFALSFTTPRDEVA--PSFGTLATVNVALNMLGBYLDKKEBTLQAVGLST 1643  
 DB 2577 -VSESNVL-----LDEVELTDPKIQALLTV----- 2601

QY 1644 QABGTRTLKSLMFTMENCFLYLLISQAMRYLRDA--VHPRDKQMKQELSELSTLISL 1702  
 DB 2602 ---LATLVKTYTTFDFDRIIYE---YLAESAVVFPKVPFVAVNHLDSKINTLLSLC 2651

QY 1703 -----SRVFRGAPSSPATGVLP-----POGSTSLSKASPSQ 1737  
 DB 2652 QDPMLNPIHIGIVQSVUYNHESPQYQTSYLSQSGFNGIMRFACRFPKQOI--PDVA 2707

QY 1738 EPLIQVQAFV 1748  
 DB 2708 ELIVKFDAL 2718

RESULT 32





Db 2495 -----PRARKMSLDMGQPSQANTKKLLGTRKSPDHLISDTK-----APKREM 2538  
 Qy 1526 PSAAASAPSSSKOPADPTASQOALHTVQYGLKILSKTALAHFTPPVCCILLDQSL 1585  
 Db 2539 ESGITTPPMRVAETDYEMETORISSSQHPHLKRV----- 2576  
 Qy 1586 DLAEYNFLPALFTPTFSEVA--PSFCTLATATNALMMLGELDKKKKPELTQAVGLST 1643  
 Db 2577 -VSESNNL-----LDEVLTPKICQALLTV----- 2601  
 Qy 1644 QAEGRFTLKSLMTMENCFTYLLISQAMRYLDP--VHPRDKQRMKQELSELSTLSSSL 1702  
 Db 2602 -----LATVKKYTTDEPDQRIIYE-----YLAASVFPKYPVAVHNLLDSKINTLLSLC 2651  
 Qy 1703 -----SRYRRGAPSSPATGVLP-----PQGSTLSKASPEEQ 1737  
 Db 2652 QDPNLLNPIHGIQGVSVVHESPPQYQTSYLSQSFQNGLMRPAGPFSKQTOI-----PDYA 2707  
 Qy 1738 EPLIQVQAFV 1748  
 Db 2708 ELIVKFLDALI 2718

RESULT 33  
 US-09-385-222A-4  
 ; Sequence 4, Application US/09385222A  
 ; Patent No. 6465211  
 ; GENERAL INFORMATION:  
 ; APPLICANT: RIKEN;  
 ; APPLICANT: Kanehiko Mihoehiba  
 ; TITLE OF INVENTION: High affinity IP3 binding polypeptide  
 ; FILE REFERENCE: PH-657US  
 ; CURRENT APPLICATION NUMBER: US/09/385,222A  
 ; PRIOR FILING DATE: 1999-08-26  
 ; PRIOR APPLICATION NUMBER: JP98/242207  
 ; NUMBER OF SEQ ID NOS: 14  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 4  
 ; LENGTH: 2749  
 ; TYPE: PRT  
 ; ORGANISM: Mus musculus  
 US-09-385-222A-4

Query Match 1.4%; Score 123; DB 4; Length 2749;  
 Best Local Similarity 18.2%; Pred. No. 0.28;  
 Matches 306; Conservative 222; Mismatches 523; Indels 630; Gaps 80;

Qy 2 IRKSKITSVLSFCSSSRELMTILLGRSALRELSQIEALNKHMRLLLEGLSYKPPSPSS 61  
 Db 610 ITAAEIDTFVSLVRKNRE-----PRELDYLSDLCVSMNKS-----IPVQEL 651  
 Qy 62 AEKYKANDVASPLKEGLRISKF-----LGIDEGSVOLLOQCYLOEDYRG 108  
 Db 652 ICKAVLNFTNADILLETGLVLSREFBSGVSTGENALEGDEEB-----VWLFMDSNKEI 707  
 Qy 109 RD-SVKTVLQDEROSQALIKIADYV-YE-----ERTC-----ILRC 143  
 Db 708 RSKSVRELDAQAKGQKEDRDILSYRYQNLFARMCLDRQYLAINEISQGLDVLDIRC 767  
 Qy 144 VLHLITTFQDERHRYRYVADCC-----VDKLEKLVS-KYRQOQEBELYKTEA-PTWE 193  
 Db 768 M-----SDENLYDYDRASFRCRLMLHMHVDRPOEQVTPVKYARLWSEIPISEIADIDYD 820  
 Qy 194 THGNLTERQVSRWFVQ-----CLR-----EQSMLLEII-----FLYYAY 228  
 Db 821 SSGG--SKDEIKERFAQTMERVEEYLDVYQCRFPFSDKEKXKLTFEVVMANRLITFGF 878  
 Qy 229 FEMAPDULVLTKNF-----KQGFGRS-QTNRHLDVDETMDPFVD 267  
 Db 879 YNF--SDLLRLTKILAILDCVHTTIFPISKMTKGEBKNGSNMRSIHGVELMTQVVL 936  
 Qy 268 RIGFISALI-----LVGMDLBSLHKCALDR--ELHQAQDGLICQDMCMLT 316

Db 937 KGGFLEPTMPAAAPBEGNVQAEBEKEDIMVMDTKITIILOPILNVRIDYRISCLICI 996  
 Qy 317 FGDIPHNAFVLLAMALLRHTLNPEETSSVVRKIGTAIQ--LNV-----FOYLRLOSL 369  
 Db 997 F-----KREPBESNQSSETSSGNSQGBPSNVPGALDFEHEBOAEGI 1040  
 Qy 370 ASGNDCTTACKCVYGLISFVLTSLHTLNQODIIDTACEVLADPELPELFWGTEP 429  
 Db 1041 FGGSE-----NTPLDLDDHGR----- 1058  
 Qy 430 TSGGIITLDSVCGMFPHLSPLOL-----RATVSGK--STAKVYS 470  
 Db 1059 -TFLRVLIHTMDYRPLVSGALQLEFRHSQROEVLQAFKQVQLVTSQVDNYKOIQ 1117  
 Qy 471 FLDMKSFY--NELYKXK-----PHD-----VISHEGTLW----- 498  
 Db 1118 DLDDLRISIVKSELWVYKGGPDEPMDGASGENHKKTEGTSKPLKHSTSSYNRVVK 1177  
 Qy 499 -----RROTPKLPLPGGQTNLRIPQTVGOYMLDRAVLRWEY 538  
 Db 1178 EIIIRLSKLCVQESASVRSRKOORLBRMGANA-----VLE-----LQIDPY 1222  
 Qy 539 SYSWTLFTCEIEMLLHVSTADVIOH-CORVAPIIDLVKVISTDLSADCLPITSRI 597  
 Db 1223 EKADETMQ--EIKRLAH-----EFLQFCAQNOONALHKHINLPKPGILEAVTMQHI 1276  
 Qy 598 YMLLQRLTVISPPVDIVASCNCLTVLAARNPAKWTDLRHGFLPEFVAP--VSSL 653  
 Db 1277 FMNNFOLCSINE--RVQHFHICIEF-HGRN-----VQYIKFLQTVKABGKFKKC 1326  
 Qy 654 SOMISAEQMGAG-----GYGNL--LMSQPOQS-----YGVIAIRLRITTL 694  
 Db 1327 QDMWAEVLVNSGDLVFNDRASFOTLIQMKSERDRMDESNPLMHHILVELAVCTE 1386  
 Qy 695 VKQLGSTQSGLVF-----CWMFV-----LKEMLPZYHK 724  
 Db 1387 GKAVYTEIKCNSLPLPDDIVRVVTHEDCIPVXIAVINFLNHCYVDTEVEMKEIYSNMH 1446  
 Qy 725 WRNSHGVRQICLIIEILHAILNICHETDLHSHTPSLQFLCTGSLAYTEAGQTVINI 784  
 Db 1447 WK-----LENFVLVDICRACNNTS--DRKADS-----ILEKYVTEIYMSIVT- 1487  
 Qy 785 MGIGVDITDMVMAOPRSDQAG--QOGOLLIKTVLAFSVTNVNRILKPPSVVSPLEQ 863  
 Db 1488 -----TFSSPFSQSTLTQTRQVYFVOLLQGVRYH-----CWLMPISQK 1529  
 Qy 844 ALSQGHAGNNLIYAKYIYHKDPALPR-LAIQ-----LKRLATVAPMSVYA 892  
 Db 1530 ASVE-----SCIRVLS-----DVAKSRALAIPLVLDQVNNLFLKSHNIYQKTALMW 1576  
 Qy 893 CIGNDAAAIRDAFLT-----RLO--SKIEEMRKIMLEFLTYAVET--QPGV- 936  
 Db 1577 RLSARNAARDSDVLAASRDYRNIETRLDVISALSD-RLRPLVQAELSVLVVDLHREPLL 1635  
 Qy 937 -----IEFLNLEVDKQSGSKFSL 957  
 Db 1636 FPEPTDARKCBESGGRICIKLKHQQLLENEKLCIKYLOTRKEMTMDRGEGEQIIS 1695  
 Qy 958 GMSCLHAYL-----ELIDSOQODRYWCPPLHLRAALIAFLHALWODRRDSAMLVLRKPF 1013  
 Db 1696 D--ESENAEIPOAPEAENSTEOLEBSP-----LROLEDHKGKGAALQILVNRYY 1744  
 Qy 1014 W-----ENLTSPFGLTSP--PSET--SEPSILETCLIMKIIICLEIYVYVKGSL 1059  
 Db 1745 GNIRPSGRRESLSFENGPLSPGGPCKGGGCGPSSSTSGENSLAEVQCHLDKEGAS 1804  
 Qy 1060 DOSLKDTLKKFSIEKFAWSGVYSLAVHVAETBSGCTSLLEYQMLVSNARMMLLIAT 1119  
 Db 1805 NLVI-DLINNASSDRFH-----ESILAIALLBSGNT-----I 1838  
 Qy 1120 THADIMHLDTSVVRROLPLVDIDGTAL--LIVPASVNCRLGSKCTLLILILRQWR 1176

Db 1839 QHSFECRLTEBCKSEKFEKVFYDRMKVAQOEIKATVNTSDLGKCKD-----DEVDR 1892  
Qy 1177 ELGSDVDELGLPTELEBGLADQOMKTKA--VF----- 1211  
Db 1893 DAPSKRKAKEPTQITE--EVRDQLLEASATRAFTTFRREADPDHYOGSECTQATT 1949  
Qy 1212 -----SAFTYVLO-----MKEMKVSIDIPOYSQOLVYN-----VCETLO----- 1243  
Db 1950 DKAKODLEMSAVITIMQPIKLPQLCENHRDLQNFRCQNKKNVNLVCEITLOFLUCI 2009  
Qy 1244 -----EEVIALFDQTRHSLALGSAETEDKDSMETDCCSRHRDRQDQVC 1287  
Db 2010 CGSTTGGLGLGLYNEKVALINQTLSTL-----TEYCOGCPCHENQN--C 2053  
Qy 1288 V 1288  
Db 2054 I 2054

RESULT 34  
US-09-228-986-78  
Sequence 78, Application US/09228986  
Patent No. 6359198  
GENERAL INFORMATION:  
APPLICANT: Strabala, Timothy  
APPLICANT: Nieuwenhuizen, Niels  
TITLE OF INVENTION: Compositions Isolated from Plant Cells  
TITLE OF INVENTION: and their Use in the Modification of Plant Cell Signalling  
FILE REFERENCE: 11000/1020  
CURRENT APPLICATION NUMBER: US/09/228,986  
CURRENT FILING DATE: 1999-01-12  
NUMBER OF SEQ ID NOS: 130  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 78  
LENGTH: 903  
TYPE: PRT  
ORGANISM: Eucalyptus grandis  
US-09-228-986-78

Query Match 1.4%; Score 122; DB 3; Length 903;  
Best Local Similarity 22.6%; Pred. No. 0.042;  
Matches 111; Conservative 66; Mismatches 168; Indels 146; Gaps 26;  
Qy 395 SLEHTLTGNO-QDIIDTAC-EVLADPSL--PELFWGTEPTSG---LGIIIDSYCGM--FP 445  
Db 507 SLEVTSGNLCLSFSTACGDASSPSIEAPQVITVPERNKGNHRLAIILGAVGVSLA 566  
Qy 446 HLLSPLOL-----RALVSGKS-TAKKVSFLDKMSFYNELYKHKPHDYS 491  
Db 567 ILILPLFVFMYRRRGTETMSYTERAVADVNMNAKIFSYKEIKATNNF-----KEYIG 621  
Qy 492 HED-GTLWRQTPKLLYPLGGQTNLRIPQTVGVMLDRAVLVWEXSYSSWTLFTCEI 550  
Db 622 HGSFSGSYLGNLP-----VGKLV-----AVKRAFDTQGLGADSFINEV 659  
Qy 551 EMLLHV-----VSTADVIQHCORVKEPIIDLVHKVISTDLSTADCLPITSRIYML--LQR 603  
Db 660 RLLSGVRHQNVLVSLEGFYESQRQ---ILVEYELPG-SLADQLYGNSRKFSLSWVR 714  
Qy 604 LTTVISPVDVYIASCVCNCLTVLAARNPAKWTDLRHTGFLPFVAHPVSLQMSIAEGMN 663  
Db 715 LK-----IADDAKG-----LDVLIHNSNPRRIIHRDIKCSNIILDKEMN 753  
Qy 664 AG--GYGNLNMNSEOPQGEYGVTTAFLRLITTLVKGQSGTOSQGLVPCWVFLKEMLP 721  
Db 754 ARLCDFG--LSKOMIQPDATH-----VTTVVGKTAGYLD-----PE 787  
Qy 722 YHKRYSNHSVREQ-----IGCLILELITHALINLCHETDLSHSTPSIQFLCISLAYTE 776  
Db 788 Y-----YSTQQLTEKSDVYSFGVVLLELI-----CGREPLNHSGRPDSFNLVLMKPYLQ 837  
Qy 777 AGQTVINIMIGVDTIDVWMAAQRPSDGAEGOGQOLIKTVKLAFSVTNNVIRLKP--S 835

Db 838 AGAFEIVDESIG-----GSFVDSMRKVAKIAVRSVERDASIRPTIA 879  
Qy 836 NVSPLEQALS 846  
Db 880 QILSVLKEAYS 890

RESULT 35  
US-10-101-464A-78  
Sequence 78, Application US/10101464A  
Patent No. 6768041  
GENERAL INFORMATION:  
APPLICANT: Strabala, Timothy  
APPLICANT: Nieuwenhuizen, Nicolaas  
APPLICANT: Higgins, Colleen M.  
TITLE OF INVENTION: Compositions Isolated from Plant Cells  
TITLE OF INVENTION: and their Use in the Modification of Plant Cell Signalling  
FILE REFERENCE: 11000.1020C2  
CURRENT APPLICATION NUMBER: US/10/101,464A  
CURRENT FILING DATE: 2002-03-18  
PRIOR APPLICATION NUMBER: 09/704,302  
PRIOR FILING DATE: 2000-11-01  
PRIOR APPLICATION NUMBER: 09/228,986  
PRIOR FILING DATE: 1999-01-12  
PRIOR APPLICATION NUMBER: 60/162,866  
PRIOR FILING DATE: 1999-11-01  
PRIOR APPLICATION NUMBER: PCT/US00/00724  
PRIOR FILING DATE: 2000-01-11  
NUMBER OF SEQ ID NOS: 989  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 78  
LENGTH: 903  
TYPE: PRT  
ORGANISM: Eucalyptus grandis  
US-10-101-464A-78

Query Match 1.4%; Score 122; DB 4; Length 903;  
Best Local Similarity 22.6%; Pred. No. 0.042;  
Matches 111; Conservative 66; Mismatches 168; Indels 146; Gaps 26;  
Qy 395 SLEHTLTGNO-QDIIDTAC-EVLADPSL--PELFWGTEPTSG---LGIIIDSYCGM--FP 445  
Db 507 SLEVTSGNLCLSFSTACGDASSPSIEAPQVITVPERNKGNHRLAIILGAVGVSLA 566  
Qy 446 HLLSPLOL-----RALVSGKS-TAKKVSFLDKMSFYNELYKHKPHDYS 491  
Db 567 ILILPLFVFMYRRRGTETMSYTERAVADVNMNAKIFSYKEIKATNNF-----KEYIG 621  
Qy 492 HED-GTLWRQTPKLLYPLGGQTNLRIPQTVGVMLDRAVLVWEXSYSSWTLFTCEI 550  
Db 622 HGSFSGSYLGNLP-----VGKLV-----AVKRAFDTQGLGADSFINEV 659  
Qy 551 EMLLHV-----VSTADVIQHCORVKEPIIDLVHKVISTDLSTADCLPITSRIYML--LQR 603  
Db 660 RLLSGVRHQNVLVSLEGFYESQRQ---ILVEYELPG-SLADQLYGNSRKFSLSWVR 714  
Qy 604 LTTVISPVDVYIASCVCNCLTVLAARNPAKWTDLRHTGFLPFVAHPVSLQMSIAEGMN 663  
Db 715 LK-----IADDAKG-----LDVLIHNSNPRRIIHRDIKCSNIILDKEMN 753  
Qy 664 AG--GYGNLNMNSEOPQGEYGVTTAFLRLITTLVKGQSGTOSQGLVPCWVFLKEMLP 721  
Db 754 ARLCDFG--LSKOMIQPDATH-----VTTVVGKTAGYLD-----PE 787  
Qy 722 YHKRYSNHSVREQ-----IGCLILELITHALINLCHETDLSHSTPSIQFLCISLAYTE 776  
Db 788 Y-----YSTQQLTEKSDVYSFGVVLLELI-----CGREPLNHSGRPDSFNLVLMKPYLQ 837  
Qy 777 AGQTVINIMIGVDTIDVWMAAQRPSDGAEGOGQOLIKTVKLAFSVTNNVIRLKP--S 835  
Db 838 AGAFEIVDESIG-----GSFVDSMRKVAKIAVRSVERDASIRPTIA 879  
Qy 836 NVSPLEQALS 846

```

Db      880 QILSVLKEAYS 890
      :|:|:|
RESULT 36
US-10-028-946-4
; Sequence 4, Application US/10028946
; Patent No. 6734009
; GENERAL INFORMATION:
; APPLICANT: Yu, Xuanchuan
; APPLICANT: Miranda, Maricar
; TITLE OF INVENTION: No. 6734009e1 Human Kinases and Polynucleotides Encoding the Same
; FILE REFERENCE: LEX-0289-USA
; CURRENT APPLICATION NUMBER: US/10/028,946
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: US 60/258,335
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1958
; TYPE: prt
; ORGANISM: homo sapiens
US-10-028-946-4

Query Match      1.4%; Score 122; DB 4; Length 1958;
Best Local Similarity 18.1%; Pred. No. 0.18;
Matches 274; Conservative 226; Mismatches 544; Indels 468; Gaps 67;

Qy      441 GCMFPHLSPILQILRALVSG-----KSTAKKVSFLDKMSFYNELYHKRPHDIVSHEDG 495
Db      105 CGHFHE-----YQVVRKATGDIYAMKMKKALLAQGVSEFFEE-----ERNILSRSTS 154
Qy      496 TLMRRQETKLTPPLGGQINRL-----PGTVGQVMDLDRAYLVIRMEYSYSMTL--FTC 548
Db      155 P-W--ITQLQAFODKNHLYLVMETOPG-----DLGLNLYEDQDENTLIQFYLA 203
Qy      549 EIEMLHVVSTADVIQHCORVPE--IIDLVHKVISTDLIA-----DCLLPITSR 596
Db      204 ELILAVSHVLMGVYH--RDIKENIIVDRGHILVDFGSAKKNKMNATLPICTP 261
Qy      597 IYMLQRLTIVYISP-----PVDVASCVNCLTVLAARNPAKVTDLRHTEGFLP 644
Db      262 DYMAPEVLTIVNKGCGKTYGLDCDWMVSVGLA--YEMTIGRSP----- 302
Qy      645 FVAHPVSLSQMISAEKNGAGYGNL-----LNNSEPOGEYGTIAFLILTTIVYQG 698
Db      303 -----FAGTSAKTFNNINMFORFLKFPDDPK-----VSDFLDLIQSLGCG 345
Qy      699 LGSQTQOGLVPCVMEFVLKEMLPYHKKRYSNGVAREQIGCLILIELHAILMLCHETDLHS 758
Db      346 KERLKEEGL--CCH-----PPFSKIDMN--IRNSPPF-----VPLTKSDDDTSN 387
Qy      759 SHTSLSQLCLCSLAYTAGQTVINIMIGIVDTIDMMAAPRSDGAEQCGCG----- 811
Db      388 FDEBEKN-----SWSSSPQOLSPSGFSGEBELPFVGFYSKALGILGRSESVVSGIDS 440
Qy      812 -----QLIKTVKLAFS-----VTNNVRLKPPSNV 838
Db      441 PAKTSSMEKKLILSKELQDSODCKHKEQEMTLHRVSEVAVLISQKVELLAASETOR 500
Qy      839 SPLEQALSQHGAGNNLAVL--AKYIVHKDPAFLRLAQLRLATVAPMSVYACIGN 896
Db      501 SILRQDLATYITECSSLSKRLSEQARMEVSGEDDK--ALQLLDIR-----E 544
Qy      897 DAAAIRDAFLRLRQSKIEDMRIKMILIEFLVAVATQGLIELPLNEVGDGSGSKEPS 956
Db      545 QSRKIQETKEQEQAYQVEEMRLMNQLEEDIVSARRRDLYE--SELRSRLAAAEFK 600
Qy      957 LGMWSCULH-----AVLELDSQOQDRYH--CPPLIRHRAIYFLHA--LMQ 997
Db      601 RKATECQHKLKADQKPEVGEYAKLEKINAEQOLKIQELQETLEKAVKASTATELLQ 660

```

```

Qy      998 DRDQSMVLVTRKPKFWENLSPFLGTLSPSPSESESILET-----CALIMKICILEIYY 1053
Db      661 NTRQKERAERELEKLNREDS-----SEGIRKKLVAEERRSHLENKVRLE-- 708
Qy      1054 VVKGSILDOSLKDITLKFSIEKRFAYMSGYKSLAVHVAETGSSCTSLLEYQMLVSARM 1113
Db      709 -TMRERNRLKDDIQTSQO-----IQNADKLELEBK--REMQ----- 746
Qy      1114 LLIATTHADIMHUTDSVVRQLFDVLDGFKALLLVASVNCRLGSMKCTLLILLRQ 1173
Db      747 ---VSAQHLLEV-HIKQKQOHEEKIKVLD-----NQ 773
Qy      1174 MKRELGVDELGLPTEILHGVQADQOLMEKTKAKVS--AFITVLOMK-----EMKVS 1226
Db      774 IKQDLAD-----KETTENWQRRHEEZHNE--KGKILSEKATINANDSKINSLEQRIV 824
Qy      1227 DIPQSQVLNVG-----ETLQEVIALPDQTHSIALGASATEDKDSMETDD--CSRSRH 1279
Db      825 ELSEANKLANSSLFTQRNKKAOEEMISELRQOKFYLETQAGKLEAQRKLEBQLEKISH 884
Qy      1280 RDQRDGVGVGLHIAKELCEVD--EDGSMVQVTRRLPILPTLLTLEVSILRMKON---L 1334
Db      885 QDHSQKRL--LELETRLEVSLEHEBQKLELRQ-----LTELQSLQERESQUTAL 935
Qy      1335 HFTFATLHLTLTARQOAGATAVAGAGITQSICLPLSYQLVSTNGTAQTPSARSKSLDA 1394
Db      936 QANRAALESQRLQAKTELETTAEAEETIAL-----TANHDEIQRK--FDA 980
Qy      1395 PSMPGVYRLSMS--LMEQLKTLRYN-----FLPALDFV--GVHQ----- 1431
Db      981 -----LRNSCTVITDLEQNLQTEDNAELNQNPFYLSKQLEBASGANDIEIVQLREVD 1034
Qy      1432 -----ERTLOCINAVRTYQSL--AC-----LEBADHTVGGIILQSLN 1465
Db      1035 HLRREITEREMQLTQSKQKMEALKTTCTMLEBOYMDIEALNDELLEKEREQWEAMRSVLGD 1094
Qy      1466 FMKEMHFLPOLMRDQVNLGYLQACTSLHSRKMLOHYLQNNCGD--GLPSAVAQRYQ 1523
Db      1095 EKSQFEBEVRVRELQMDTEKQSRARADQRTESRQVVELAVKHKHKEIILALQALKEQKL 1154
Qy      1524 RPPSASAPSSSKQRA--ADTASBQALHTVOYGLKTLISKTALALRHFTPDVCOILL 1581
Db      1155 KASISDKLANDLEKKNALMEKNARSLOQKLETERELKQRLIEE-----QAKL 1201
Qy      1582 DOSLDLAENVFLPALSTPTTPDSEVAPSRCTILATVNNVALMNGELDKKEPITQAVGL 1641
Db      1202 QQQWDL-QKNHITRLT-----QGLQEBALDRADLDLKTERSDL 1236
Qy      1642 STQAEGRTRLKSLIMFTMENCFYLLISQAMR--YLDPVAVHPRDKOR----- 1686
Db      1237 EYQLENIQVLYSHEKYVMEG--TISQOKLIDPQAKMDQPAKKKVAVLYQNELKTLAL 1292
Qy      1687 -----MKQELSELSTLSSLSRYFRGGA--PSSPAT-----GVLPSPQG 1724
Db      1293 EKERKARCAELMELQKTRIELRSAREBAARAKATDHPHPTPAYARQAIMSAIVRSPH 1352
Qy      1725 KSTSLSKASPS 1736
Db      1353 QPSAMSLIAPPS 1364

RESULT 37
US-10-028-946-2
; Sequence 2, Application US/10028946
; Patent No. 6734009
; GENERAL INFORMATION:
; APPLICANT: Yu, Xuanchuan
; APPLICANT: Miranda, Maricar
; TITLE OF INVENTION: No. 6734009e1 Human Kinases and Polynucleotides Encoding the Same
; FILE REFERENCE: LEX-0289-USA
; CURRENT APPLICATION NUMBER: US/10/028,946

```

CURRENT FILING DATE: 2001-12-20  
 PRIOR APPLICATION NUMBER: US 60/258,335  
 PRIOR FILING DATE: 2000-12-27  
 NUMBER OF SEQ ID NOS: 4  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO 2  
 LENGTH: 2054  
 TYPE: PRT  
 ORGANISM: homo sapiens  
 US-10-028-946-2

Query Match 1.4%; Score 122; DB 4; Length 2054;  
 Best Local Similarity 18.1%; Pred. No. 0.2;  
 Matches 274; Conservative 226; Mismatches 544; Indels 468; Gaps 67;

441 CGMPHLLSPILLQILRALVSG-----KSTAKKVSFLDKMSFYNELYHKPHDIVSHEDG 495  
 105 CGHFAE-----VQVVRKATGDIYAMKWKKKALLAQGVSFEE-----ERNILSRSTS 154  
 496 TLMRQRPKLYPLGGQTNLRI-----PGTVGVMLDDRAVLVMEYSYSMTL--FTC 548  
 155 P-W---IPQLQYAFQDKNHLVLMVEYQPG-----DLSLNKYEDQDLDELITQFYLA 203  
 549 EIEMLHVSADYVIOHCQVYP---IIDLVKIVSTDLSIA-----DCLLPITSR 596  
 204 ELILAVSHVLMGVYH--RDIKPENIIVDRGHIKLVDPGSAAKNSKMNNAKLPICTP 261  
 597 IYMLQRLTVVISP-----PVDVIASCNCLTVLAARNPAKVWTDLRHTGFLP 644  
 262 DYMAPEVLTVMNCGDKGTGLDCDWMVSGVIA---YEMITYGRSP----- 302  
 645 FVAIPVSLSQMISAEKNAGGYGNL-----LMSBQPGEGYGTIAFLILITTVYQG 698  
 303 -----FABGTSARTENNINMFORFLKFPDPC---VSSDFLIDLSLTCGQ 345  
 699 LGSTOGVLVPCVMEVFLKEMLPYHKMRVNSHGVREQICLLIELIHAIINI,CHETDLAS 758  
 346 KERKLFEBL--CCH-----PFEKIDMNN--INSPRP-----VPLKSDDDTSN 387  
 759 SHTSLOFLCISLAYTAGQTVINIMIGIVDTIDMAAQRSDGAGGCGG----- 811  
 388 FDEBEKN-----SWSSSPQQLSPSGFSGEELPFVGFYSKALGIGRSESVSGDS 440  
 812 -----QILITVKLAFS-----VTMNVIRLKRPSNV 838  
 441 PAKTSMEKLLIKSKELQDSQDKCKHMOEWTLRHREVEAVLSCKEVELKASETOR 500  
 839 SPLFOALSQHGHNLI,AVL--AKYIVKHDPALPLAIDLLKRLATVAPMSVYACIGN 896  
 501 SLBEDLATTYTTCCSLKRSLEQAMEVSOEDDK---ALQULHDIR-----E 544  
 897 DAALIRDAFLTRLOSIEDMRIKWMILFELTVAVETOPGLIELFLMLEKQSDSGSKEFS 956  
 545 QSRKQEIKEQEQVQAVEEMRLMNQLEBDLVASARRSDVY---SLRSRLAAEFK 600  
 957 LGMNSCLH-----AVLELDSQOQDRYV--CPPLHRAALAFHA--LWQ 997  
 601 RKATCCQHLKAKDQGRVEGEYVKLRINABOOLKQELQEKLEKXVASTETTELQ 660  
 998 DRDSAMLVLRTRKPFMENTLSPFGTSPSESTSEPSILET---CALIMKIICLEY 1053  
 661 NIRQAKERAERLEQLQNRDS-----SEGIRKGLVAREERHLENKVRLE--- 708  
 1054 VVKSGLDLSLDKTLKKSIEKRFAYVSGVYSLAVVAETGSSCTSLLEYQMLVSAWM 1113  
 709 -TMRERENRLKDDITKSCQ-----IQOMADKILELEBKH---REAQ----- 746  
 1114 LLIATTTADTLMHLLDSVVRRLFLVDYDGTALLLVASVNCRLSGMKCTLLILLRQ 1173  
 747 ---VSAQHLLEV-HLKQEQHYEEKIKVID-----NQ 773  
 1174 WKREIGSVDELIGPLTEILEGLVADQOOLMEKTKAVFS---AFITVLQMK---EMKVS 1226

774 IKKOLAD-----KETLEMMQHRHEBAHE-KGILISEQKAMINAMDSKIRSLREQIV 824  
 1227 DIPQYSQVLAVNC-----ETLQSEVIALPQUTHSLALSGATBDKSMETDD--CSNRH 1279  
 825 ELSANKLANANSFLTORNMKAQOEMISELROQFYELETQGLKEADNRKLEOLEKISH 884  
 1280 RDQRDGVCVGLHHLAKELCEVD--EDGSMLOVTRRLPIPLTLTTLEVSIMKON---L 1334  
 885 QDHSQDKRL--LELETRLREVSLEHEOKLELKQ-----LELOLSLOERSQULTAL 935  
 1335 HFTETATLHLLTLARTOQATVAVAGITQSIGCPPLSVYQSTNGTAQTPSARKSIDA 1394  
 936 QAAABALESQORQKTELEETTAEBEERIQUL-----TAHRDEIQRK--FDA 980  
 1395 PSWPGVTRLSMS---LMEQLKTLRYN-----FLPEALDPV--GVHQ----- 1431  
 981 -----LNSCTVITDDEBQNLQTEBDVAELNQNFIYSKQLDASGANDIYOLRSEVD 1034  
 1432 -----ERTLOCANAVRTVQSL--AC-----LEADHTVGFILQLSN 1465  
 1035 HLRREITEREKQLTQOKOTMEALKTTCMLBEQVMDLEALNDELLEKROHEAMRSVLGD 1094  
 1466 FMKEWHPLPOLMEDIQVNLQYLOACTSLHSRKMLOHYLQNRNGD--GLPSAVQORVQ 1523  
 1095 EKSQFECGRVRELQMLDTERKQSRABDQRTESQVVELAVKEHKAETIALQOLKEQKL 1154  
 1524 RPPSAASAPESSSQPA--ADTEASBOQALHTVOYGLKILSKTIALRHFTPVQOILL 1581  
 1155 KAESISDKLNDLEKHMALENNARSLOQKLETERELKORLEB-----QAKL 1201  
 1582 DQSLDLAEVNFPLFSFTTTPFDEBVAVPSFGTLATVAVNALMGEIDPKKEPITQAVGL 1641  
 1202 QOQMDL--QKHIFILT-----OGLOALDRADLLTERBDL 1236  
 1642 STQAEGRITLKLMTFMCNYLLISQAMR--YLBDPAVPRDKOR----- 1686  
 1237 EYQLENIQVLYSHKRVKMEG---TISQOTLIDFLOAKMDQPAKKKVPLOYNELKLAL 1292  
 1687 -----MKQELSESLTSLSSRYFRGA-----PSPPAT-----GVLPSQOG 1724  
 1293 EKERACAELEBAQKRIEILRSARBEAMARKADPHPSPTATARQOIJAMSAIVRSPER 1352  
 1725 KSTSLKASPEP 1736  
 1353 QPSAMSLAPPS 1364

RESULT 38  
 US-08-952-127-12  
 ; Sequence 12, Application US/08952127  
 ; Patent No. 6211336  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Shilo, Yosef  
 ; APPLICANT: Tagle, Danilo A.  
 ; APPLICANT: Collins, Francis S.  
 ; TITLE OF INVENTION: ATRAXIA-TELANGIECTASIA GENE  
 ; NUMBER OF SEQUENCES: 24  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Kohn & Associates  
 ; STREET: 30500 No. 6211336thwestern Hwy., Suite 410  
 ; CITY: Farmington Hills  
 ; STATE: Michigan  
 ; COUNTRY: U.S.  
 ; ZIP: 48334  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30  
 ; CURRENT APPLICATION DATA:  
 ; FILING DATE: US/08/952,127  
 ; CLASSIFICATION: 530

```

ATTORNEY/AGENT INFORMATION:
NAME: Kohu, Kenneth I.
REGISTRATION NUMBER: 30,995
REFERENCE/DOCKET NUMBER: 2290.00029
TELECOMMUNICATION INFORMATION:
TELEPHONE: 810-539-5050
TELEFAX: 810-539-5050
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 3066 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Mus musculus
US-08-952-127-12

Query Match 1.4%; Score 122; DB 3; Length 3066;
Best Local Similarity 17.8%; Pred. No. 0.43;
Matches 298; Conservative 234; Mismatches 552; Indels 592; Gaps 81;

QY 89 DEEGSV--QLQCVIAGEYRGRDVKIVYLODERGSAIILKIDYEEETCLIRCVLH 146
DB 1269 DEKSIANOIQKCM-----KSLVD-----CEPKLIYH 1296
QY 147 LITVFQDE--RHRYVEYADCVKLEKELVSKYRQOFELKYTEAPTWETHNIMTERQV 204
DB 1297 ILPFAVAGTDSY-----VSQKRETKAKYVD-----LKGEDFLKQI 1335
QY 205 SRMFQCCRBOGSMLEITFLYAYAFEMAPSDILVITKMFKEQGRSGRTNHLVDET--M 262
DB 1336 DQVFSNLP--IVVELMTLHETADSADSD-----ASQATALCDFSGDL 1379
QY 263 DPFVDRIGYFSAIILVEGMIDIESLHKCALDDRBELHGFADGILIQDMDCMLTFGDIR 321
DB 1380 DPAPRPFPSPVIVIOATFAYISNCKTKYKSLIEI-----LSKIPD 1420
QY 322 HHAPVLLAMALLRHTLNPEETSSVVRKIGTAIQLVNQ--YLITLQSLASGNDCTTST 380
DB 1421 SYQKILLIIC-----EQAAETNNVFK--HRIKIYHLPVSLIKDIOGSGAMAFV 1471
QY 381 AOCMCYGLSFV-----LTSLHITLGNODIIDTAGEVLADSLPELFGTEPTGGLG 434
DB 1472 LRDIYITLHYINKSSHFTDVSLSFSLCDDLSRVCHTAVTQCKAL-----ESH 1525
QY 435 IILDVSCGM--FPHLSPLQLLRALVSGKS-----TAKVYSPFLDKMSFYNEL--- 481
DB 1526 VIVGTLPLVDYQVEQGVLDLKLIVIDKNKNLSTITKLDLPFDHVIY--KDLRLTQ 1584
QY 482 ----YKAKPHDIVISHEDGT--WRQTPKLLYPLGQQTNLRIPGTGVQVMLDRAVYLRW 536
DB 1585 QKIRYSGGFSILREINHFLSVSAVNPFLRLTEGLKDLRLRQLRQHKDMQD---LIRA 1640
QY 537 EYSSSWTLFPCIEMLHVVSTA-----DIVQHC--QRVKPI-----IDLVHVIS 581
DB 1641 SQNDPQDGIIVKLVLSLQTSKMAVNOTGEREVLEAVGRCLGELPDESTIAVQ--N 1697
QY 582 TDISIADCL--LPTSRIYMLQRLT---TVISPPVDVASCVCL--TVLAARNPAKWT 635
DB 1698 KDVAITYKAYGPEBDERELQWTLIMLTALNNTLVEDSVKIRSAACLNILATIKIGHFE 1757
QY 636 DLR-----HTGFLPFAVAPVSSLSQMTISABGNAGGYGNLNMSEQDGE-- 680
DB 1758 NYKTSADPMTLYLOPFRFSRKPFLEVPKVEKVL--EGDLA-----VNLWMPQESH 1808
QY 681 ----YGVTIAFLRLITLVKQLSGTOSQGLVPCVMFLKEMLPYHKMYNKHGVAEQI 736
DB 1809 DIWKITITCAFL-----DSGGINSTIQLKRWCE-----VATDF 1843
QY 737 GCLILE-LIH-AIINLCHET--DLHSHSTPSLOPLCISLAYTEAGQTVININGIDVTI 792
DB 1844 COMLLPYLIHDVLLQDTHESWRTLLSAVRG--FFTSQFHKSSQASR----- 1888

```

```

QY 793 DMWMAQPRSDGAGGOGGULLI--KTVKIAFSTVNNVIRKPPSN----- 836
DB 1889 ----SATPANDSESENFELRCCLDKKSQRTMLAVDYLRRQKSSGTADPAFWLIDNY 1944
QY 837 -VVSPLRQALSOH-----GAGNNLLIYAKYIYHKH 867
DB 1945 LEVAKVAGSCSAHTLALVAEIYSDKSTDEQEKSPTEEGSGTT--ISLSKSEET 2003
QY 868 DPALPRLAIQLKRLATVAPMSYACIGNDAAIRDAFLTRLOS--KIDMRIKYMIIEFL 926
DB 2004 GISLQDILILEIYSIG--EPDSLYGCGGKWLQ--LIRIRYEHEATWEKALVYDL 2057
QY 927 TVAVET---QPGILEFLNLEVVDGSDGSKPEFLGMMSCAHVLELDSQOQRYWCPPL 983
DB 2058 ETSISSSTROSGIIOALQ--LGSILHSIYVYKGLDYER--REWCAL 2101
QY 984 LH-PAATAFLHALW-----QDRRDSAMLVLRK--PKFMENTSPLEFGL 1025
DB 2102 QELRYQAMNMQGLCASAGQVEGTSYHESLYNMLQCLRNREBFTEYSLR---YASL 2158
QY 1026 SPSPSESEPSILETCALIMKITCLEIYVYVKSGLDQ--SIKDTLKRFST----- 1072
DB 2159 FRVKEVEELS-----KGSLESYYSIYPTLSRLQALGELNSELFL 2198
QY 1073 ----EKRPAYW-----SGYKSLAVHVAETGSSGTS- 1100
DB 2199 SRSTVDREREBAAYWKQKHSQILKDDPSFOEPLMLRTYILETLVQKEMERGQACSKD 2258
QY 1101 ----LLEYOMLVSAW-----RMLLIATTHADI-----MHITDSV--RRQLFL 1138
DB 2259 ILTKHLVEFSLARTFNQPLPERALFKIQYNSALGISEMHLBEAQVWAKQESLAL 2318
QY 1139 DVDL-----GFKALLVPASVNCRL--GSM-----KCTLLILILR 1172
DB 2319 SILKQMIKKIDSSFQKENDAGLKVY--AECLRVGSGWLAETCLENPVIMQTYLE 2373
QY 1173 QMRBELGSDLEILGPTEILEGLVADQOLMEKTKAKVSAFITYVMKEMKVSIDIPOY- 1231
DB 2374 KAVKAVAGSYD--GNSRELNGQMK-----APLSLARFPDQYQRIENY 2415
QY 1232 ----SOLVNVCTTLOBEVIALFDQTRHSLALGAT--EDKDSMETDCSRSHRDR 1283
DB 2416 KSEPFENKQTL--KRAKEVGLL--REHKIQTRRYVVKQRELEDECAIRALREDR 2469
QY 1284 DGVCVLGHLAKELGCV-----DEGDSWLOVTRRLPLPLTLTLEVSLMKON 1333
DB 2470 ----KRLCAVENYINCLSGEHDLM--VFRLCSLMLENSGVSEVNGMKKD 2517
QY 1334 LHFEATLHLTLTARTQOAGATAVAGAGITQSGIPLSVYQLSSTN--GTAQTPSASRSL 1392
DB 2518 ----GMIKISSYKFLPLM--YQAAAMGTRMTGG----- 2544
QY 1393 DAPSWPGVYRLMSLMELKTLIRYFLPEALDPVGVQHOERITLOCLNAVRYQSLACLE 1452
DB 2545 ----LGFHB-----VANNISIRSLDH 2562
QY 1453 ADHTVGFILQLSNFMKEMHFLPOLMRDIOVNLGY-----LCOACTSLHS 1498
DB 2563 PHHTLFIILANANKEFLSKPRTTRRSRIRKTSKSHSHDEDRTEAATKIHS 2618

RESULT 39
US-08-328-254-6
Sequence 6, Application US/08328254
Patent No. 5710022
GENERAL INFORMATION:
APPLICANT: Zhu, Xueliang
APPLICANT: Lee, Wen-Hwa
TITLE OF INVENTION: A No. 5710022el Nuclear Mitotic Phosphoprotein
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Campbell and Flores

```

STREET: 4370 La Jolla Village Drive, Suite 700  
 CITY: San Diego  
 STATE: California  
 COUNTRY: USA  
 ZIP: 92122  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/328,254  
 FILING DATE: 24-OCT-1994  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/141,239  
 FILING DATE: 22-OCT-1993  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Campbell, Cathryn A.  
 REGISTRATION NUMBER: 31,815  
 REFERENCE/DOCKET NUMBER: P-CJ 1191  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (619) 535-9001  
 TELEFAX: (619) 535-8949  
 INFORMATION FOR SEQ ID NO: 6:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2482 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-328-254-6

Query Match 1.3%; Score 121; DB 1; Length 2482;

Best Local Similarity 18.5%; Pred. No. 0.36;

Matches 395; Conservative 329; Mismatches 743; Indels 668; Gaps 105;

18 RELWTLLGRSALRELQIEALNKGWRRLLEGLSYKPPSPSAEKVKANKDVAPLKE 77  
 48 RLNLNVLDSKS-----VEVETQK-----LAYMELQCKAKESDDQKHQKEI-----EN 88  
 78 LGLRISKPLGL-DEEGSVOLL-----OCY-LOEDYRGTRDSVK-----TYVLQDERQ 121  
 89 MCLKTSQUTGVDELEHKLQLLSNEIMDKRCYQDLHAEYSRLRLKSKDASLVTNBDH 148  
 122 SQALILKADYYERTCILKCVLHL--TYFQDERPRYVE-----YADCYDK 168  
 149 QRSILA-----FDQOPMHHSFANIIGEGSMPSERSECRLEADQSPKNSAILQNRVDS 202  
 169 LE-----KELVSKYKROOFELYTEAPTWETHGNLMTERGVSWMFV-----QCLRE- 214  
 203 LEFSLESQKQMSDLQKCEBELVQIKG--EIEENMLKAEQMHQSFWAETQSRIKSLQED 259  
 215 -----QSMLEIIFLYAAFEWAPSDLVLT-KMFEKQG-FGSRCQNRHLVDETMDPFYDR 268  
 260 TSAQNVAATL-----SALENKEKELQLNDKVEEQAIEQLKKSNNLLEBSL-----KE 311  
 269 IGYSAIALLVEGMDIESLHKALDRRELHOPADQGLICQDMDCMLTFGDIPIHAPVLL 328  
 312 LQLSETSLERKEKMSIISL--NKREIEBELTQNGTKLEINA----- 352  
 329 AMALLRHLNPEETSVVRIKGTALQNLN-----VFOYLTR--LLOSILASGN-- 374  
 353 -----SLN-OEKKNLLOKSESFPANYIDEREKSISELSDDQKELIILQRCSETGNAY 404  
 375 -----DCTTSTACM--CYUGLISFVLTSLELHTLGNQODIIDTACEVLA--D 417  
 405 EDLSQKYAAQAEKNSKLECLNLC-----TSL-----CENRKELEQLKEAPAKHQ 451  
 418 PSLPELFGTEPTSGLIILDSVCGMFRHLSPLLIQILRALV-----SGKSTA----- 465  
 452 EFLTGLAAEERNQMLLELFTV-----QALRSEWTDNQNNKSKSBAAGIKOEI 500  
 466 -----KKVVSFLDKNSFYNELYKHKPHVDIYSHBDGLYMRQRPKLLLYPLGGQTN 514

501 MTLKEQNKKQKENVNDLLQENBQMLKVMKTK-HECQMLESEPI--RNSVXERESEBRQCN 557  
 515 LRIPQG--TYGVWLD-RAVLWMEYSYSWTL-----FTCEIMLHVST--ADV--- 562  
 558 FK-PQMDLEVAEISLDVYNQVLQLEAMLRKKEIKQSEKEXKCLHELTQINGDLETS 616  
 563 -----IHCORVVRITDLYHKVTS-----TDLIADCLLPITRIYMLL 601  
 617 NLQMGQOEISGLKQCE-----IDAEKYISGFHELSQNDNHNQCSLOTMMKLNEL 671  
 602 QRLTTVISPVDVASCVN-----CLTVL--AARNPAKWTDLRHTGFLPFVAHPVSSLS 654  
 672 EKICEIILQAEKYELVTELNDNSRSECITATRMABEVGLINEVK-----ILNDSGLL 724  
 655 QMISAEQNMAGYCNLNMSEQPOGEYGVITA-----FLRLITLVKQGLSGTOSQ 705  
 725 HGEIVEDI PGGEFG-----EQPNEQHPVSLAPIDESNSYEHLLTSDKEVQMHFAELQEK 778  
 706 GLVPCWFLVKEMLPYHKMRVNSH-----GVREQIGCL 739  
 779 FL-----SLQSEHKLIHDQCOMSSMSELQTYVDSLKAMLVLTNLNRFQDL 828  
 740 ILLELIHAILNCHETDLSHSTPLQFLCI-----CSLAYTEAGQTVINIMIGVDTID 793  
 829 VKEM-----QLGIEBGLVPSLSSCVDDSSLSLSDGSSPYRLLEQTG-----D 873  
 794 MVMQAQPSDGAEGOGQQLIKTVKLAFSVTNNVIRLKPSPNVVSPLEQALSGHAGN 853  
 874 MSLLS--NLBGA--VSANQCSVDDEVFCSLSQIEENLTKETPSAPAKGVEELSECEVVRQ 929  
 854 NLIAVLAKYIYHKHDPRLPRLAQLKRLANVAPMSVYACIGNAAAIROAFRLRLOSKI 913  
 930 SLEKLEEM--ESQGITKKEKIEQLERQLS--SERQELDLCKRKYLSNENQOQKLSVT 985  
 914 EDKRIKWLIEFLT--VAVETQPGIIEP-----FLNLVNDVDSGSKSEFSLGMMWC 962  
 986 LEMSKLAABEKQEQEQLSLELVARLQJGLDLSRSLSLIGTDEDAIGRNE-----SC 1039  
 963 -----LNAVLELID-----SQOQRIW 979  
 1040 DISKEHTSETTERPPKPDVHQCIDKDAQDNLMDIEKITEGALKPTGECGSEBSPDTNY 1099  
 980 CRP-----LHRAATLFLHAL-----WODRBSAMLVTRKPKFMENTL 1017  
 1100 EPPGEDKTQSSBECISLSESGFNALVPMDFLQNOEDIHNLQLVKETSNNELILHYIE 1159  
 1018 --TSPLEGTLSPPSEETSEPSILETCALIMKI-ICLAIYVVVGSLL--DQSLKDTLKKS 1071  
 1160 DDRKRVESLNMKELEDSKHLQVQVQMTKIEACIELEKIV-GEKKENSPLSEKLEVFS 1218  
 1072 IE-----KRFAYWGSYVKSIAVH-----VAETBGSCTSLLEYOMLVSANRML 1114  
 1219 CDHQLLQRVATSEGLNSDLEMHADKSREDIGDNVAKVNDXWYERFLVDVNELSR 1274  
 1115 LIATTHADIMHLDVSVVRQLPLVDLDTGKALL-----LVPASVN 1155  
 1275 --ISEKASIEH--EALVLEADLEVQTEFLCLEKONENKQKIVCLLEBELSVVTSERN 1329  
 1156 CLR-----LGSWKCTLLILLRQWK--RELGS----- 1180  
 1330 QLRGELDTMSKKTALQOLSKMKEXKQELSHQSECLHCTQVABAEVXKTELLQTLSS 1389  
 1181 -VDEILGLPTEILRGV--LQADQQLMEKTKAVFSAPITVLQMTKMYSD-----LPQ 1230  
 1390 DVSELKDKTKHLOKQSLERKDSQALSTTKCELNOIQAQLKKEKELLVYKSEBSIQARISE 1449  
 1231 YSQVLNVCETLQOEIVIALPQTRHSLALGSAITEDKQSMETDDGSRSHRQORQVCLG 1230  
 1450 SDYEKLANSKALE--AALVKEGEPALRLSTQEBVHQLAR-GIEKLKVRLEADB--KKQ 1503  
 1291 LHLAKELCEVDEDDGSW-----LOV--RLRPIL-----PRLITLLE-- 1325

```

Db      1504 LHIAEKLERERENDSLKQVENLEBELQMSSENOELVILDAENSKAEVETLKQIEMMA 1563
Qy      1326 -----VSLRM-KONTL-HFTE-----ATLHLLTLANT-----QOGA 1354
Db      1564 RSLKVFELDLVTLRSEKENLTKQIOEKQGLSELKTLSSFKSLLEKEKQAEIQIKESK 1623
Qy      1355 TAVAGAGITQSIICPLLSVYQSLTNGTAQTPASRSKSIDAP-----SMP 1398
Db      1624 TAVE---MLQN-QLEIMNAVAALCGDQIMKATQSDPRLEERHOLNRIEKLRALE 1679
Qy      1399 GYRLSMSIMEQLKTLRYNPLPEALDVGVOHERTQCLNAVRTVQSLACLEADHTVG 1458
Db      1680 ADEKQQLCVLQOLKESEH-----ADLLKGRVENLEBELIARTNQHAAL-EAENSKG 1732
Qy      1459 FILQSNFMKEWHFLLPOLMRDIOVNLGLCOACTSLHSRMLQHYQNKNGDGLPBAV 1518
Db      1733 EVETLKAKIE---GHTQSLRGLBDV-----VTIRSEKENLTNELOKQO----- 1773
Qy      1519 AQRVQRPSPASABSPS---SSKQPADTEASEQALHTVQYGLKILSKTLAALRHFPD 1575
Db      1774 -ERISELEIINSFENILQEKQEKVKQMKESSTAMENLQ-TQLEINERVAAL-HNDQE 1830
Qy      1576 VC-----QILLDQSLDLAEVNFALSTPTPTPDSVAPSEFGTLLA 1616
Db      1831 ACKAKEONLSQVECLELEKAQL--QGIDEAKNNYI-----VLQS 1869
Qy      1617 TVNVALNMLGE---LDKKEPRLTOAVGLSTQAEGRTRLKSLMTMENCYLLISQMR 1672
Db      1870 SVNGILQIEVEDQKQLEKKEDEISR---LKNQIQOEQVLSKL-QVGEHQLMKEONLE 1925
Qy      1673 YLRDPAVHRDKQRMKQELSELSTLLSLSLRYFR 1707
Db      1926 -LRNLTVLELQKIQVLQSKNNAILODTLEVLQSSYK 1959

RESULT 40
US-08-970-269A-29
; Sequence 29, Application US/08970269A
; Patent No. 5976803
; GENERAL INFORMATION:
; APPLICANT: Kathryn Meek
; TITLE OF INVENTION: Genetic Test For Equine Severe
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dr. Benjamin A. Adler
; STREET: 8011 Candle Lane
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: Apple
; OPERATING SYSTEM: Macintosh
; SOFTWARE: Microsoft Word for Macintosh
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/970, 269A
; FILING DATE: No. 5976803ember 14, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Adler Ph.D., Benjamin A.
; REGISTRATION NUMBER: 35,423
; REFERENCE/DOCKET NUMBER: D5860
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 713-777-2321
; TELEFAX: 713-777-6908
; INFORMATION FOR SEQ ID NO: 29:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2987 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear

```

```

; MOLECULE TYPE: protein
; HYPOTHETICAL: no
; ANTI-SENSE: no
; ORIGINAL SOURCE:
; FEATURE:
; US-08-970-269A-29

Query Match      1.38; Score 120.5; DB 2; Length 2987;
Best Local Similarity 17.68; Pred. No. 0.57;
Matches 387; Conservative 294; Mismatches 711; Indels 749; Gaps 99;

Qy      4 KSKITS-----VLSFC-----RSRRLMTLLRSRLR----- 31
Db      32 KSWMTSTVRBPKLPVLAGCLKGLSLMNCNTKSMNEEDPQTSKEIFDAL-KAIRPIDL 89
Qy      32 -----ELSQIEABLNKMRRLLEGLSYKPPSPSSAEKYANKNDVAPLK 76
Db      90 KRYAVPLAGLCPLFLNHSQFSTCLLENYVSLFEVLSTKMGCHTINIELKKA-AHSALBSFLK 148
Qy      77 ELGLRISKPIGLDEEQQVQLQCYLOEDYRGRTDSVTVQDERQOALLIKTADYYEE 136
Db      149 QVSEFWAK---DAERHKKLYQFMEOFYGIIRNM-----DSNSKDSIALRGYGLFA 197
Qy      137 RTCLLRCVLHLLTYFODERHRYREYVADCYDKLEKELVSKYRQOFELYTEAPYETHG 196
Db      198 GPC-----KVINAKQVDPMYVELI---QRCKQLPILQIDTVDDHI 234
Qy      197 NMTHEROVSNFPOCLREBSMLLEIT-----FLYAVFENAPSDLLV 238
Db      235 YQMS-----FLQSVSVLYLDITPEVYTPVLEHLMVQIDSPPOYSP-KMQPVCCRA 287
Qy      239 LTKMFK-----BOGFGSRQTNRLHVDFT----- 261
Db      288 IVKLFLALAKGVLMNCISTVHGLIRCSKPVVPQKAGSSESDYHSEARFGKMK 347
Qy      262 MDPFVDRIGYSALLIVEGM-----DIESLHKCALDRRLHQPADGL- 305
Db      348 MPYKDYLDLFRYLLSCDQMDLSLADAEFLFVNSSLHSLNRLLYD---EFVSVLK 401
Qy      306 IQCDMDCIM--LRFQGIPIHAPVLLW-----ALRTINBERNS----- 343
Db      402 IVKLDLTLEKQNVGQEDTEATGVVPIPTSDPANLHPARKKDSAFINLVEFCREIL 461
Qy      344 -----SVARKIGTAIQ-LNVQVYLRLLQSLASGG 373
Db      462 PEKHVEFEPFWVVSFAVELILQSTRPLLSVFKLSVAVRANKMKYEGVQPKQKXS 521
Qy      374 NDCITSTACMCYVGLIS-----FVLTSELTGNGQOIIDTACEVLADPSLP 421
Db      522 PEDLEKYSQCFALFAKSKSEVSIKMQYKDELLASCITFLISLPHDIIEIDVRAVYV-PALQ 580
Qy      422 ELF---WGTEPNSGLGI-IIDSVGC-MFPHLSPLIQ-----LLAALVSGKSTAKK 467
Db      581 MAFKGLSTPLAEVGLNABEMSGYTCCKHVLOPYKYDILPSLDGLTKTSVSDETKSN 640
Qy      468 VYSEFLDKMS---FYNELYGH-KPHDVISHEDGTL--WRQTPKLYPLPGGQTNLRIPQ 520
Db      641 QVSLARAAQKGNKVKVLLHGLTKTSISSNALSLEEVARIYVRILGSLGGQNNKLVYA 700
Qy      521 TVGQVWLD---DRAVLYRMWYSYSW-----TLFTGCI 550
Db      701 ASSDEMKKCVAMDRKRLRAVPEFMKPVLYLDLFLPRVTELASASDRQTYAACBL 760
Qy      551 --EMLHVSTA-DVLOHCORVAPIIDLVHXYISTDLSIADCLPITSRIY--MLIQ--- 602
Db      761 LHSNWMFMKGATQMEEDQSGSPMQLYKRIFFVILRLACVDVDTROLVLEPLWQLIH 820
Qy      603 -----RLTVISPPVDVIAS---CVNCLTVLAANPAKWTDLNHTGF 642
Db      821 WFTNNKFEQSDTVALLFTILDGIVDPVDSITLRDFGCGIQEFLK-----WS-IKQTFP 873
Qy      643 LRFVVAHPVS--SLSQMISAGMAGGYGNULMNSQPOSEYGVTTAFLRLITTVYKQGLG 700

```



```

Db      874 000EKS P V N T S L P F R L Y S F A L H P N A F R K L ----- G A S I A F N N I Y R E F ----- 916
Qy      701 S T O S G I V P C W F V L K E M L P S Y H K R R V N S H G V R E O I G ----- C I L E I L I A I L 748
Db      917 - R E E S S L V E Q P V F --- E A L V T Y M E S L A L A H T D E K S L G T T O C C D A I D H L S L I E K H S L 972
Qy      749 N L C H E T D L S H S T P S L Q F I C I S L A Y T A G C T V I N I M I G V D T T I D M W A A O P R S D G A G Q 808
Db      973 N K A K R R L P R G P P A T S - L C L ----- L D V Q W L L A N C R P T E C R H 1012
Qy      809 G O G O L L I K T V ----- K L A F S V T N N V I R L K P S N V S P L E Q A L S Q H A G N N L A V L A K - 861
Db      1013 K S I E L F Y F V T L L P G N K S P F L M K D I I K K E D I S F L I N T F E ----- G G S G R P S G I L A Q P 1066
Qy      862 Y I Y H N D A L P R I A Q L K R L A T V A P M S V V A C L G D A A I R D A F I T R Q S K - I E M R K V 920
Db      1067 T L F H L O G F S L R A A L Q M M D M L ----- L A A L E C Y N T F I E K T L E A P K - 1108
Qy      921 M I L E F L V A V E T O P G L --- I E L F L - N L E V K O S D G S K E F S L G M S C L A V L E L I D S Q O D 976
Db      1109 ----- L G T E T O S L M K A V A F L E S I A M H D I M A E K Y F G T G ----- A T G N R P S P Q E G E 1155
Qy      977 R Y --- W C P P L L R A A I A F L A L M O D R D S A M L V I R T K P F M E ----- N L T S P L F G T 1024
Db      1156 R Y N S K C T I V R - I M E F T T L ----- L S T S P E G W K L E K O V C N T N L M K L A V K T 1202
Qy      1025 L S P S E T S ----- E R S I L E T C A L I M K I C L E I Y Y V K S L D O S L K D T L K K S I E 1073
Db      1203 L C E P S I G F N I G D V A V M N Y L P S V --- C T N L M K A L K K S P Y --- K D I L E M L K E K I T A O S I E 1256
Qy      1074 K R P A V S G V K S L A V H A V A E T E G S C T S L E Y O M L V S A M R L ----- L I A T T H A D I M H 1126
Db      1257 E L C A ----- V D I X C P D --- A C V D R A R L A S V S A C Q L H R A G V L C V I I P S O S A D O H N 1304
Qy      1127 L T D S V R R Q L P L D V L D G T K A L L V P A S V N C L R G S M K C T L L I L R Q W K R E I G S V D E I L G 1186
Db      1305 S I G T L L S L V Y K S I A P G D E Q C L P S L D P N C K R L A G L E L A F ----- A F G 1349
Qy      1187 P L T E L L E V L Q A D O O L M E K T A K V S A P I T V L Q W ----- K E M K V S D I P O Y S Q L V I N V C E T 1241
Db      1350 G L C E H L V S L L ----- L D T V L S M P S R G S O K N I V S F S H G E Y F Y S L F S E T 1393
Qy      1242 L Q E V I A L F D Q T R H S I A L G S A T E D K D S M E T ----- D D C S R S R H R D Q D G V C V L G L H A 1294
Db      1394 I N T E L L K N L D L A --- V L E L M K S V D N P K V S N V L N G M L D O S P R D R S E K H Q ----- G L K L A 1446
Qy      1295 K E L C E V D E D G D S W ----- L Q V T R R L P ----- I L P T L L T L E V S L 1328
Db      1447 T I I L Q N M K K C D S W M A K O S A P E S K M A V L T L A K I F O I D S V C F N T N H C M F P E V F T T Y - V S L 1505
Qy      1329 - - R M K O N H F E A T I L L T L A R T O G A T A V A G A I T O - S I C L P - - L S V I Q L S T N --- 1379
Db      1506 L A D S R L D L H L K G O A I L L P F F ----- T S L G S L E D L K V L E N L I V S N P M K S E B R P P 1558
Qy      1380 G T A O T P S --- A S R K S L D A P S M P G V Y R L S M S --- I M E O L K T L R Y N F L P E A L D F V G V H Q 1431
Db      1559 G T L Q V N N V D C M K K F L D A --- L E L S X S P M L Q L M T E I L C R E O Q H M E B L P O S T F K K I 1612
Qy      1432 E R T L O C I N A V R T V O S I A C L E A D H T V G F I L O L S N F M K E ----- W H F H L P O L --- 1477
Db      1613 A R K S C I T O L G L E S V Y M F R R D D L --- L S N I T R O A F V D R S L L T L L M H C S L A N A L R E F 1666
Qy      1478 ----- M D I Q V N L G Y L C O A C T S L S H S R K M L O H Y I Q N K G D G L P S A V A Q R V O R 1524
Db      1667 F S K I V E A I N V L K S R F I K L N E S A F D T O I T K M G Y Y K M L D ----- 1705
Qy      1525 P P S A S A A P S S S K O P A D T E A S E O A L H T V O Y G L L --- K I L S K T I A L R H - - F T P D V C - 1577
Db      1706 ----- V M S R L P K D V H S K E S K - I N Q V H G S C I T E G S E L T K I L I C L Y D A F T E M N A G 1756
Qy      1578 - - Q I L L D O S L - D L A E Y N ----- F L P A L S F T T P T P D S E V A P S F G T L L A 1616
Db      1757 E N Q L E R R L Y H C A Y N C A I S V A C V F N E L K F Y O G F L F T --- E K E K N I L I F E N L I D L K R 1813

```

```

Qy      1617 T V N V A L N M L G E L D K K K E P L ----- T O A V G L S T O A E G T R T L K S L M F T M E N C F Y L L I S Q A M 1671
Db      1814 C Y T F F I E V E V M E R K K T L E I R K A R E A A A S G D S D G P R Y I S L ----- 1856
Qy      1672 R Y L R D P A V H P R D K O R M Q O E L S - E L S T L S L S R Y F R R G A P S P A T G V L P S P O G K S T S L S 1730
Db      1857 S Y L A D S ----- L S E M S Q P D S T G V O S Y S ----- Y S S Q D P K S T T A H 1893
Qy      1731 K A S P E S O E P L I Q ----- L V O A F R H M O R 1753
Db      1894 F R R O K H K E S M I Q D D I L E L E M D E L N O H C M A T M T A L I K H M O R 1934

```

Search completed: December 30, 2004, 15:15:55  
 Job time : 113.809 secs

***This Page Blank (uspto)***

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: December 30, 2004, 15:26:41 ; Search time 191 Seconds  
(without alignments)  
42.174 Million cell updates/sec

Title: US-10-719-385-20  
Perfect score: 14  
Sequence: 1 MIRKSTSVLSFC 14

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 1825181 seqs, 575374646 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Listing first 150 summaries

Database :

UniProt 02: \*  
1: uniprot\_sprot: \*  
2: uniprot\_trembl: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	8	57.1	0821R8	Q821R8 chlamydom
2	7	50.0	92 1 YB41 STRWU	P59154 streptococ
3	7	50.0	145 2 Q7Y1T8	Q7Y1T8 cryptospori
4	7	50.0	347 2 Q7N4K4	Q7N4K4 photorhabdu
5	7	50.0	431 2 Q831Y4	Q831Y4 enterococcu
6	7	50.0	447 2 Q81G21	Q81G21 bacillus ce
7	7	50.0	696 2 Q8EN85	Q8EN85 oceanobacil
8	7	50.0	987 2 Q8BN14	Q8BN14 debaryomyce
9	7	50.0	1018 2 Q8FME0	Q8FME0 corynebacte
10	7	50.0	1019 2 Q8NMS1	Q8NMS1 corynebacte
11	7	50.0	1019 2 CAFE20751	Q8NMS1 corynebacte
12	7	50.0	1027 2 Q6CIE3	Q6CIE3 kluyveromyce
13	7	50.0	1601 2 Q45231	Q45231 caenorhabdi
14	6	42.9	43 2 Q48626	Q48626 leuconostoc
15	6	42.9	56 2 Q8ZIC7	Q8ZIC7 yersinia pe
16	6	42.9	56 2 AAE63080	AAE63080 yersinia
17	6	42.9	81 2 Q9P1F0	Q9P1F0 homo sapien
18	6	42.9	81 2 P70991	P70991 bacillus su
19	6	42.9	84 2 Q8KV60	Q8KV60 bacillacteri
20	6	42.9	101 2 Q8KHV2	Q8KHV2 renibacteri
21	6	42.9	105 2 Q7QT66	Q7QT66 glardia lam
22	6	42.9	126 2 Q918L5	Q918L5 brachydanio
23	6	42.9	131 2 Q7UDV4	Q7UDV4 rhodospirill
24	6	42.9	143 1 RL15 METVA	RL15032 mechanococ
25	6	42.9	143 2 Q6LXDI	Q6LXDI mechanococ
26	6	42.9	143 2 CAFE30977	CAFE30977 methanoco
27	6	42.9	153 2 Q7VC75	Q7VC75 prochloroco
28	6	42.9	157 2 Q7RKH6	Q7RKH6 plasmodium
29	6	42.9	167 2 Q737A5	Q737A5 bacillus ce
30	6	42.9	167 2 AAE41657	AAE41657 bacillus
31	6	42.9	169 2 Q70GJ2	Q70GJ2 pyrus commu

105	6	42.9	404	2	Q8DNC7	Q8dnc7 streptococc
106	6	42.9	404	2	Q8NZL8	Q8nzb8 streptococc
107	6	42.9	404	2	Q89YB6	Q89yb6 streptococc
108	6	42.9	404	2	Q87NMG	Q87nm6 streptococc
109	6	42.9	404	2	Q7CET9	Q7cet9 streptococc
110	6	42.9	405	2	Q83SR1	Q83sr1 enterococcu
111	6	42.9	405	2	Q8DSG1	Q8dsg1 streptococc
112	6	42.9	412	2	Q879J7	Q879j7 streptococc
113	6	42.9	413	2	Q8R3P7	Q8r3p7 mus musculu
114	6	42.9	414	2	Q6L150	Q6l150 paracentrot
115	6	42.9	414	2	Q6TUG5	Q6tug5 lytechinus
116	6	42.9	429	2	Q8IC35	Q8ic35 arabidopsis
117	6	42.9	429	2	Q8SAD0	Q8sad0 arabidopsis
118	6	42.9	433	1	DHOM_BACSU	P19582 bacillus su
119	6	42.9	433	2	Q9KFE1	Q9kfe1 bacillus ha
120	6	42.9	438	2	Q8WTK2	Q8wtk2 caenorhabdi
121	6	42.9	439	2	Q8L160	Q8l160 plasmodium
122	6	42.9	447	2	Q6HLG3	Q6hlg3 bacillus th
123	6	42.9	447	2	Q7JBB0	Q7jbb0 bacillus ce
124	6	42.9	447	2	Q81T78	Q81t78 bacillus an
125	6	42.9	447	2	AAS40438	Aas40438 bacillus
126	6	42.9	447	2	AAT30506	Aat30506 bacillus
127	6	42.9	450	2	Q7RRR1	Q7rrr1 plasmodium
128	6	42.9	450	2	Q7V3E3	Q7v3e3 prochloroco
129	6	42.9	452	2	Q8SGX8	Q8sgx8 arabidopsis
130	6	42.9	469	2	Q8TUV5	Q8tuv5 methanopyru
131	6	42.9	469	1	NDOU_BUCA1	P57264 buchiera ap
132	6	42.9	479	2	Q6C5S9	Q6c5s9 yarrowia li
133	6	42.9	479	2	Q8BLE6	Q8ble6 mus musculu
134	6	42.9	482	2	Q80527	Q80527 arabidopsis
135	6	42.9	490	2	Q7RXK3	Q7rxk3 neurospora
136	6	42.9	493	2	Q8BTP7	Q8btp7 debaryomyce
137	6	42.9	519	2	Q982D1	Q982d1 rhizobium 1
138	6	42.9	519	2	Q9UPX4	Q9upx4 homo sapien
139	6	42.9	526	2	Q9NUF1	Q9nuf1 caenorhabdi
140	6	42.9	530	2	Q6R6M4	Q6r6m4 homo sapien
141	6	42.9	530	2	AAR91701	Aar91701 homo sepi
142	6	42.9	539	2	Q9RDS9	Q9rds9 streptomyce
143	6	42.9	540	1	GRBE_HUMAN	Q14449 homo sapien
144	6	42.9	552	2	Q6NKL7	Q6nkl7 corynebacte
145	6	42.9	552	2	CAE48512	Ca48512 corynebact
146	6	42.9	558	2	Q8BMQ4	Q8bm4 mus musculu
147	6	42.9	559	2	Q6CHS9	Q6chs9 yarrowia li
148	6	42.9	560	2	Q82235	Q82235 arabidopsis
149	6	42.9	561	2	Q961V1	Q961v1 drosophila
150	6	42.9	566	1	SNFC_YEAST	P53628 saccharomyc

## ALIGNMENTS

RESULT 1

ID	Q821R8	PRELIMINARY:	PRT:	506 AA.
AC	Q821R8:			
DT	01-JUN-2003 (TrEMBLrel. 24, Created)			
DT	01-JUN-2003 (TrEMBLrel. 24, Last sequence update)			
DT	01-MAR-2004 (TrEMBLrel. 26, Last annotation update)			
DE	60 kDa chaperonin, putative.			
GN	OrderedLocuNames=CCA00870;			
OS	Chlamydomonas reinhardtii			
OC	Bacteria; Chlamydiales; Chlamydiales; Chlamydiales; Chlamydiales			
OX	NCBI_TaxID=83557;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=GPIC;			
RX	MEDLINE=22569155; PubMed=12682364;			
RA	Read T.D., Myers G.S.A., Brubaker R.C., Nelson W.C., Paulsen I.T.,			
RA	Heidelberg U.F., Holtzapfel E.K., Knouri H.M., Federova N.B.,			
RA	Carty H.A., Umayam U.A., Haft D.H., Peterson J.D., Beanan M.J.,			
RA	White O., Salzberg S.L., Hsia R.-C., McClarty G., Rank R.G.,			
RA	Bavoli P.M., Fraser C.M.,			
RT	"Genome sequence of Chlamydomonas reinhardtii (Chlamydomonas reinhardtii) GPIC):			

RT examining the role of niche-specific genes in the evolution of the  
 RT Chlamydiales.  
 RT Nucleic Acids Res. 31:2134-2147(2003).  
 CC -1- FUNCTION: Prevents misfolding and promotes the refolding and  
 CC proper assembly of unfolded polypeptides generated under stress  
 CC conditions (By similarity).  
 CC -1- SUBUNIT: Oligomer of 14 subunits composed of two stacked rings of  
 CC 7 subunits (By similarity).  
 CC -1- SIMILARITY: Belongs to the chaperonin (HSP60) family.  
 DR EMBL: A016997; A05611.1; -.  
 DR HSP60; P45746; ISRV.  
 DR TIGR: CCA00870; -.  
 DR GO: GO:0005524; F:ATP binding; IEA.  
 DR GO: GO:0003754; F:chaperone activity; IEA.  
 DR GO: GO:0006457; P:protein folding; IEA.  
 DR InterPro: IPR001844; Chaperonin Cpn60.  
 DR InterPro: IPR002423; Cpn60/TCP-1.  
 DR InterPro: IPR008950; GroEL-ATPase.  
 DR Pfam: PF00118; Cpn60\_TCP1; 1.  
 DR PRINTS: PR00298; CHAPERONIN60.  
 DR ATP-binding; Chaperone; Complete proteome.  
 KW SEQUENCE 506 AA; 55976 MW; A57144199770239A CRC64;

Query Match 57.1%; Score 8; DB 2; Length 506;  
 Best Local Similarity 100.0%; Pred. No. 5.2;  
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 KITSVLSF 13  
 Db 213 KITSVLSF 220

RESULT 2

ID	YB41_STRMU	STANDARD:	PRT:	92 AA.
AC	PS9154:			
DT	28-FEB-2003 (Rel. 41, Created)			
DT	28-FEB-2003 (Rel. 41, Last sequence update)			
DT	01-OCT-2004 (Rel. 45, Last annotation update)			
DE	Hypothetical UPF0223 protein SMU.1141c.			
GN	OrderedLocuNames=SMU.1141c;			
OS	Streptococcus mutans.			
OC	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;			
OC	Streptococcus.			
OX	NCBI_TaxID=1309;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=UAI59 / ATCC 700610 / Serotype C;			
RX	MEDLINE=2225063; PubMed=12397186; DOI=10.1073/pnas.175601299;			
RA	Ajdic D.J., McShan W.M., McLaughlin R.E., Savic G., Chang J.,			
RA	Carson M.B., Primeaux C., Tian R., Kenton S., Jia H.G., Lan S.P.,			
RA	Qian Y., Li S., Zhu H., Najjar F.Z., Lai H., White J., Roe B.A.,			
RA	Ferretti J.J.;			
RT	"Genome sequence of Streptococcus mutans UAI59, a cariogenic dental			
RT	pathogen.";			
RU	Proc. Natl. Acad. Sci. U.S.A. 99:14434-14439(2002).			
CC	-1- SIMILARITY: Belongs to the UPF0223 family.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).			
DR	EMBL: A014951; AAN5883.1; -.			
DR	HMAP: MF_01041; -; 1.			
DR	InterPro: IPR007920; UPF0223.			
DR	Pfam: PF05256; UPF0223; 1.			
KW	Complete proteome; Hypothetical protein.			
SO	SEQUENCE 92 AA; 10622 MW; B8C844A96FER0BD1 CRC64;			

Query Match 50.0%; Score 7; DB 1; Length 92;  
 Best Local Similarity 100.0%; Pred. No. 13;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 ITSVLSF 13  
 |||||  
 18 ITSVLSF 24

## RESULT 3

ID Q7YY18 PRELIMINARY; PRT; 145 AA.

AC Q7YY18; 01-OCT-2003 (TREMBlrel. 25, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)  
 DE 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 GN ORFNames=1MB.86;  
 OS Cryptosporidium parvum.  
 OC Eukaryota; Alveolata; Apicomplexa; Coccidia; Bimeriida;  
 CC Cryptosporidiidae; Cryptosporidium.  
 NCBI\_TaxID=5807;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RA Bankier A.T., Spriggs H.F., Partmann B., Konfortov B.A., Madera M.,  
 RA Vogel C., Teichmann S.A., Ivens A., Dear P.H.;  
 RT "Integrated mapping, chromosomal sequencing and sequence analysis of  
 RT Cryptosporidium parvum."  
 RL Genome Res. 0:0-0(2003).  
 DR EMBL; BX538350; CAD98670.1; -.  
 KM Hypothetical protein.  
 FT NOW TRR 145 145

SO SEQUENCE 145 AA; 16639 MM; 36C9CF68CE2B298 CRC64;

Query Match 50.0%; Score 7; DB 2; Length 145;  
 Best Local Similarity 100.0%; Pred. No. 19;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 5 SKITSVL 11  
 |||||  
 27 SKITSVL 33

## RESULT 4

ID Q7NAK4 PRELIMINARY; PRT; 347 AA.

AC Q7NAK4; 01-MAR-2004 (TREMBlrel. 26, Created)  
 DT 01-MAR-2004 (TREMBlrel. 26, Last sequence update)  
 DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)  
 DE Similar to unknown protein Y14H of Escherichia coli.  
 GN OrderedLocustNames=plu2327;  
 OS Photorhabdus luminescens (subsp. laumondii).  
 CC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 CC Enterobacteriaceae; Photorhabdus.  
 NCBI\_TaxID=141679;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=TT01;  
 RX MEDLINE=22957627; PubMed=14528314;  
 RA Duchaud E., Rusnok C., Frangeul L., Buchrieser C., Gavaudan A.,  
 RA Taouit S., Bocs S., Bouraux-Bude C., Chandler M., Charles J.-F.,  
 RA Daaga E., Derose R., Derzelle S., Freysinet G., Gaudault S.,  
 RA Medigue C., Lanois A., Powell K., Sigulier P., Vincent R., Wingate V.,  
 RA Zouine M., Glaser P., Boemare N., Danchin A., Kunst F.;  
 RT "The genome sequence of the entomopathogenic bacterium Photorhabdus  
 RT luminescens."  
 RL Nat. Biotechnol. 21:1307-1313(2003).  
 DR EMBL; BX571866; CAE14620.1; -.

DR GO; GO:0016747; F:transferase activity, transferring groups O. .; IEA.  
 DR InterPro; IPR002656; Acyl\_transf\_3.  
 DR Pfam; PF01757; Acyl\_transf\_3; 1.

KW Complete proteome.  
 SO SEQUENCE 347 AA; 40078 MM; EB4C7F7429C3C299 CRC64;  
 Query Match 50.0%; Score 7; DB 2; Length 347;  
 Best Local Similarity 100.0%; Pred. No. 39;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 RKSKITS 9  
 |||||  
 336 RKSKITS 342

## RESULT 5

ID Q831Y4 PRELIMINARY; PRT; 431 AA.

AC Q831Y4; 01-JUN-2003 (TREMBlrel. 24, Created)  
 DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)  
 DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)  
 DE Adenylosuccinate lyase.  
 GN Name=purB; OrderedLocustNames=EF2361;  
 OS Enterococcus faecalis (Streptococcus faecalis).  
 CC Bacteria; Firmicutes; Lactobacilliales; Enterococcaceae; Enterococcus.  
 NCBI\_TaxID=1351;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN=V563 / ATCC 700802;  
 RX MEDLINE=22550857; PubMed=12663927; DOI=10.1126/science.1080613;  
 RA Paulsen I.T., Banerjee L., Myers G.S.A., Nelson K.E., Seshadri R.,  
 RA Read T.D., Fouts D.E., Eisen J.A., Gill S.R., Heidelberg J.F.,  
 RA Tettelin H., Dodson R.J., Umayam L.A., Brinkac L.M., Beanan M.J.,  
 RA Dougherty S.C., Deboy R.T., Durkin S.A., Kolonay J.F., Madupu R.,  
 RA Nelson W.C., Vamathevan J.J., Tran B., Upton J., Hansen T., Shetty J.,  
 RA Khouli H.M., Ueberback T.R., Radune D., Ketchum K.A., Dougherty B.A.,  
 RA Fraser C.M.;  
 RT "Role of mobile DNA in the evolution of vancomycin-resistant  
 RT Enterococcus faecalis."  
 RL Science 299:2071-2074(2003).  
 DR EMBL; AE016954; AA082083.1; -.

DR HSSP; Q9X010; IC3C.  
 DR TIGR; EF2361; -.  
 DR GO; GO:0004018; F:adenylosuccinate lyase activity; IEA.  
 DR GO; GO:0016829; F:lyase activity; IEA.  
 DR GO; GO:0009152; P:purine ribonucleotide biosynthesis; IEA.  
 DR InterPro; IPR003031; D\_crystallin.  
 DR InterPro; IPR003032; Fumarate\_lyase.  
 DR InterPro; IPR008948; L-Asparaginase-like.  
 DR Pfam; PF00206; Lyase 1; 1.  
 DR PRINTS; PR00149; FUMARATELYASE.  
 DR TIGRFAMs; TIGR00928; purB; 1.  
 DR PROSITE; PS00163; FUMARATE\_LYASES; 1.  
 KW Complete proteome; Lyase.  
 SO SEQUENCE 431 AA; 49171 MM; 5F60A97EA10F781A CRC64;

Query Match 50.0%; Score 7; DB 2; Length 431;  
 Best Local Similarity 100.0%; Pred. No. 47;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 KITSVLS 12  
 |||||  
 398 KITSVLS 404

## RESULT 6

ID Q81G21 PRELIMINARY; PRT; 447 AA.

AC Q81G21; 01-JUN-2003 (TREMBlrel. 24, Created)  
 DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE Potassium uptake protein KtrB.

```

GN ORFNames=BC13390;
OS Bacillus cereus (strain ATCC 14579 / DSM 31).
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OK NCBI_TaxID=226900;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=22608415; PubMed=12721630; DOI=10.1038/nature01582;
RA Ivanova N., Sorokin A., Anderson I., Galleron N., Candelon B.,
RA Kapral V., Bhattacharya A., Reznik G., Mikhailova N., Lapidus A.,
RA Chu L., Mazur M., Goldsman E., Larsen N., D'Souza M., Walunas T.,
RA Grechkin Y., Busch G., Haselkorn R., Fomstein M., Ehrlich S.D.,
RA Overbeek R., Kyriades N.C.;
RT "Genome sequence of Bacillus cereus and comparative analysis with
RT Bacillus anthracis."
RU Nature 423:87-91(2003).
DR EMBL, AF017002; AAP08372.1; -.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0008324; F:cation transporter activity; IEA.
DR GO; GO:0006812; P:cation transport; IEA.
DR InterPro; IPR003445; Cat transport.
DR Pfam; PF02386; TtkH; 1.
DR TIGRFAMs; TIGR00933; 2a8, 1.
SQ SEQUENCE 447 AA; 49036 MW; AA80200421EB1199 CRC64;

Query Match
Best Local Similarity 100.0%; Score 7; DB 2; Length 447;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 7 ITSUSF 13
Db 325 ITSUSF 331

RESULT 7
ID Q8EN85 PRELIMINARY; PRT; 696 AA.
AC Q8EN85;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Transcriptional antiterminaltor (BglG family).
GN OrderedLocusNames=OB2602;
OS Oceanobacillus theysensis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Oceanobacillus.
OX NCBI_TaxID=182710;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=HTE831;
RA MEDLINE=22220767; PubMed=12235376;
RA Takami H., Takaki Y., Uchiyama I.;
RT "Genome sequence of Oceanobacillus theysensis isolated from the Iheya
RT Ridge and its unexpected adaptive capabilities to extreme
RT environments."
RU Nucleic Acids Res. 30:3927-3935(2002).
DR EMBL, AP004601; BAC14558.1; -.
DR GO; GO:0003723; F:RNA binding; IEA.
DR GO; GO:0005351; F:sugar porter activity; IEA.
DR GO; GO:0003700; F:transcription factor activity; IEA.
DR GO; GO:0009401; P:phosphoenolpyruvate-dependent sugar phospho. .; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR001550; BglG_antitermin.
DR InterPro; IPR009059; bi_resp_regltr_C.
DR InterPro; IPR002197; HTH_Fis.
DR InterPro; IPR003501; Ptrans_IIB.
DR InterPro; IPR002178; Pts_EIIA_2.
DR InterPro; IPR003058; wing_hlx_DNA_bnd.
DR Pfam; PF00874; FRD; 2.
DR Pfam; PF02302; Pts_IIB; 1.
DR ProDom; PD001689; Pts_EIIA_2; 1.
DR TIGRFAMs; TIGR01199; HTH_fis; 1.
KM Complete proteome.

```

```

SQ SEQUENCE 696 AA; 80117 MW; SE7D8D0075115CD CRC64;

Query Match
Best Local Similarity 100.0%; Score 7; DB 2; Length 696;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 3 RKSKITS 9
Db 500 RKSKITS 506

RESULT 8
ID Q6BNL4 PRELIMINARY; PRT; 987 AA.
AC Q6BNL4;
DT 01-OCT-2004 (TrEMBLrel. 28, Created)
DT 01-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 01-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Similar to ap|P38810 Saccharomyces cerevisiae SBD5-binding protein
DE 3
GN ORFNames=DEHA021956;
OS Debaryomyces hansenii (Yeast) (Torulaspora hansenii).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Debaryomyces.
OX NCBI_TaxID=4959;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=CBS767;
RG GENOMEVIBRES;
RA Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,
RA Lafontaine I., de Montigny J., Marck C., Neuvéglise C., Talla E.,
RA Goffard N., Frangeul L., Aigle M., Anthonard V., Babour A., Barbe V.,
RA Barnay S., Blanchin S., Beckerich J.M., Bayne E., Bleykasten C.,
RA Boistrume A., Boyer J., Catolico L., Confenietti F., de Barvar A.,
RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppe A.,
RA Hantreay F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
RA Kerrest A., Kozul R., Lemaire M., Lesur I., Ma L., Muller H.,
RA Nicoud J.M., Nikoleki M., Oztas S., Ozier-Kalogeropoulos O.,
RA Pellenz S., Potier S., Richard G.F., Straub M.L., Suleau A.,
RA Suenne D., Tekala F., Weslowski-Louvel M., Westhof E., Wirth B.,
RA Zenitov-Meyer M., Zivanovic I., Bolotin-Fukuhara M., Thierry A.,
RA Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J.,
RA "Genome evolution in Yeasts."
RT Nature 430:35-44(2004).
RN (2)
RP SEQUENCE FROM N.A.
RC STRAIN=CBS767;
RA Genoscope;
RA Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL, CR382137; CAG88479.1; -.
SQ SEQUENCE 987 AA; 109971 MW; BF2CCD55806DABOC CRC64;

Query Match
Best Local Similarity 100.0%; Score 7; DB 2; Length 987;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 6 KITSUS 12
Db 461 KITSUS 467

RESULT 9
ID Q8FME0 PRELIMINARY; PRT; 1018 AA.
AC Q8FME0;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Putative Na+/H+ antiporter.
GN OrderedLocusNames=CE2567;
OS Corynebacterium efficiens.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

```

OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.  
 OX NCBI\_TaxID=152794;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Y-314;  
 RX MEDLINE=22723752; PubMed=12840036;  
 RA Nishio Y., Nakamura Y., Kawarayashi Y., Ueda Y., Kimura E.,  
 RA Sugimoto S., Matsui K., Yamagishi A., Kikuchi H., Ikeo K.,  
 RA Gotohori T.;  
 RT "Comparative complete genome sequence analysis of the amino acid  
 RT replacements responsible for the thermostability of Corynebacterium  
 RT efficiens.";  
 RL Genome Res. 13:1572-1579(2003).  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).  
 DR EMBL: AP005222; BAC01937.1; -  
 DR GO: GO:0016021; C:Integral to membrane; IEA.  
 DR GO: GO:0008137; F:NADH dehydrogenase (ubiquinone) activity; IEA.  
 DR GO: GO:0016491; F:oxidoreductase activity; IEA.  
 DR GO: GO:0042773; P:ATP synthesis coupled electron transport; IEA.  
 DR InterPro: IPR007182; MnhB.  
 DR InterPro: IPR003916; NADHbd\_oxrds.  
 DR InterPro: IPR001750; Oxidored\_q1.  
 DR InterPro: IPR001516; Oxidored\_q1\_N.  
 DR Pfam: PF04039; MnhB, 1.  
 DR Pfam: PF00361; Oxidored\_q1; 1.  
 DR Pfam: PF00662; Oxidored\_q1\_N; 1.  
 DR PRINTS: PR01434; NADHbGNASES.  
 DR Complete proteome; Transmembrane.  
 KW SEQUENCE 1018 AA; 110100 MW; A15SF9DAB9F36015 CRC64;  
 SQ  
 Query Match 50.0%; Score 7; DB 2; Length 1018;  
 Best Local Similarity 100.0%; Pred. No. 95;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 7 ITSVLSP 13  
 |||||  
 Db 139 ITSVLSF 145  
 RESULT 10  
 Q8NM51 PRELIMINARY; PRT; 1019 AA.  
 ID Q8NM51; Q8NM51; Q8NM51; Q8NM51;  
 AC Q8NM51; Q8NM51; Q8NM51; Q8NM51;  
 DT 01-OCT-2002 (TReMBLrel. 22; Created)  
 DT 01-OCT-2002 (TReMBLrel. 22; Last sequence update)  
 DT 01-OCT-2004 (TReMBLrel. 28; Last annotation update)  
 DE NADH:ubiquinone oxidoreductase subunit 5 (Chain L)/Multisubunit Na+/H+  
 DE antiporter, MnhA subunit (NADH ubiquinone oxidoreductase subunit 5  
 DE (Chain L)/Multisubunit Na+/H+ antiporter, A subunit).  
 GN Name=MnpA; OrderedLocNames=Cg12729, Cg3024;  
 OS Corynebacterium glutamicum (Brevibacterium flavum).  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.  
 OX NCBI\_TaxID=1718;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025 Nakagawa;  
 RA Nakagawa S.;  
 RT "Complete genomic sequence of Corynebacterium glutamicum ATCC 13032";  
 RL submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;  
 RX PubMed=12948626;  
 RA Kalinowski J., Bathe B., Bartels D., Bischoff N., Bott M.,  
 RA Burkovski A., Dusch N., Eggeling L., Eikmanns B.J., Gaiagalat L.,  
 RA Goessmann A., Hartmann M., Huthmacher K., Kraemer R., Linke B.,  
 RA McHardy A.C., Meyer F., Moeckel B., Pfeifferle W., Puhler A.,  
 RA Rey D.A., Rueckert C., Rupp O., Sahn H., Wendisch V.F., Wiegand I.,  
 RA Tauch A.;  
 RT "The complete Corynebacterium glutamicum ATCC 13032 genome sequence  
 RT and its impact on the production of L-aspartate-derived amino acids  
 RT and vitamins.";

RL J. Biotechnol. 104:5-25(2003).  
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).  
 DR EMBL: AP005282; BAC00123.1; -  
 DR EMBL: BX927156; CAF20751.1; -  
 DR GO: GO:0016021; C:Integral to membrane; IEA.  
 DR GO: GO:0008137; F:NADH dehydrogenase (ubiquinone) activity; IEA.  
 DR GO: GO:0016491; F:oxidoreductase activity; IEA.  
 DR GO: GO:0042773; P:ATP synthesis coupled electron transport; IEA.  
 DR InterPro: IPR007182; MnhB.  
 DR InterPro: IPR003916; NADHbd\_oxrds.  
 DR InterPro: IPR001750; Oxidored\_q1.  
 DR InterPro: IPR001516; Oxidored\_q1\_N.  
 DR Pfam: PF04039; MnhB, 1.  
 DR Pfam: PF00361; Oxidored\_q1; 1.  
 DR Pfam: PF00662; Oxidored\_q1\_N; 1.  
 DR PRINTS: PR01434; NADHbGNASES.  
 DR Complete proteome; Transmembrane; Ubiquinone.  
 KW SEQUENCE 1019 AA; 110212 MW; A7A7372E155E0E57 CRC64;  
 SQ  
 Query Match 50.0%; Score 7; DB 2; Length 1019;  
 Best Local Similarity 100.0%; Pred. No. 96;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 7 ITSVLSP 13  
 |||||  
 Db 139 ITSVLSF 145  
 RESULT 11  
 CAF20751 PRELIMINARY; PRT; 1019 AA.  
 ID CAF20751; CAF20751; CAF20751; CAF20751;  
 AC CAF20751; CAF20751; CAF20751; CAF20751;  
 DT 02-MAR-2004 (TReMBLrel. 27; Created)  
 DT 02-MAR-2004 (TReMBLrel. 27; Last sequence update)  
 DT 02-MAR-2004 (TReMBLrel. 27; Last annotation update)  
 DE NADH ubiquinone oxidoreductase subunit 5 (Chain L)/Multisubunit Na+/H+  
 DE antiporter, A subunit.  
 GN MRPA OR CG3024.  
 OS Corynebacterium glutamicum (Brevibacterium flavum).  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacterineae; Corynebacteriaceae; Corynebacterium.  
 OX NCBI\_TaxID=1718;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 13032 / DSM 20300 / NCIB 10025;  
 RX PubMed=12948626;  
 RA Kalinowski J., Bathe B., Bartels D., Bischoff N., Bott M.,  
 RA Burkovski A., Dusch N., Eggeling L., Eikmanns B.J., Gaiagalat L.,  
 RA Goessmann A., Hartmann M., Huthmacher K., Kraemer R., Linke B.,  
 RA McHardy A.C., Meyer F., Moeckel B., Pfeifferle W., Puhler A.,  
 RA Rey D.A., Rueckert C., Rupp O., Sahn H., Wendisch V.F., Wiegand I.,  
 RA Tauch A.;  
 RT "The complete Corynebacterium glutamicum ATCC 13032 genome sequence  
 RT and its impact on the production of L-aspartate-derived amino acids  
 RT and vitamins.";  
 RL J. Biotechnol. 104:5-25(2003).  
 DR EMBL: BX927156; CAF20751.1; -  
 KW Ubiquinone.  
 SQ SEQUENCE 1019 AA; 110212 MW; A7A7372E155E0E57 CRC64;  
 Query Match 50.0%; Score 7; DB 2; Length 1019;  
 Best Local Similarity 100.0%; Pred. No. 96;  
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 7 ITSVLSP 13  
 |||||  
 Db 139 ITSVLSF 145  
 RESULT 12  
 Q6CIE3 PRELIMINARY; PRT; 1027 AA.  
 ID Q6CIE3; Q6CIE3; Q6CIE3; Q6CIE3;  
 AC Q6CIE3; Q6CIE3; Q6CIE3; Q6CIE3;

DT 01-OCT-2004 (TrEMBLrel. 28, Created)  
DT 01-OCT-2004 (TrEMBLrel. 28, Last sequence update)  
DT 01-OCT-2004 (TrEMBLrel. 28, Last annotation update)  
DE Similar to egd|S0005930 Saccharomyces cerevisiae YPL009c.  
GN ORFNames=K14A0F27335g;  
OS Kluyveromyces lactis (Yeast).  
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.  
OX NCBI\_TaxID=26865;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NRRL Y-1140;  
RG GENOLEVURES;  
RA Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,  
RA Lafontaine I., de Montigny J., Marc C., Neuveglise C., Talla E.,  
RA Goffard N., Frangeul L., Aigle M., Anthouard V., Babour A., Barbe  
RA Barnay S., Blanchin S., Beckrich J.M., Beyne E., Bleykasten C.,  
RA Boissarie A., Boyer J., Catrolicco L., Confalonieri F., de Daruvar A.,  
RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Giropi A.,  
RA Hantreave F., Hennequin C., Janniaux N., Joyet P., Kachouri R.,  
RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,  
RA Nicaud J.M., Nikolski M., Oztes S., Ozler-Kalogeropoulos O.,  
RA Pellenz S., Potter S., Richard G.F., Straub M.L., Suleau A.,  
RA Swenne D., Tekela F., Wesolowski-Jouvet M., Westhof E., Wirth B.,  
RA Zeniou-Meyer M., Zivanovic I., Bolotin-Fukuhara M., Thierry A.,  
RA Bouchier C., Caudron B., Scarpelli C., Gallardin C., Weissenbach J.,  
RT Winkler P., Souciet J.L.,  
RT "Genome evolution in yeasts";  
RL Nature 430:35-44(2004).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NRRL Y-1140;  
RG Genoscope;  
RA Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.  
RL EMBL, CR382126; CAG90004.1;  
DR SEQUENCE 1027 AA; 116540 MW; 20B7811D1229354B CRC64;

Query Match	50.0%	Score 7;	DB 2;	Length 1027;
Best Local Similarity	100.0%;	Pred. No. 96;		
Matches 7; Conservative	0;	Mismatches	0;	Gaps 0;

QY	3	RKSKITS	9
Db	497	RKSKITS	503

RESULT 13	
045231	
ID	PRELIMINARY;
045231	PRT; 1601 AA
045231	

DT 01-JUN-1998 (TEMBLrel.. 06, Created)  
 DT 01-AUG-1998 (TEMBLrel.. 07, Last sequence update)  
 DT 01-MAR-2004 (TEMBLrel.. 26, Last annotation update)  
 DE Hypothetical protein C01B9.1.  
 GN Name=C01B9.1;  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae; Pelodierinae; Caenorhabditis.  
 OC NCBI\_TaxId=6239;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RC MEDLINE=99069613; PubMed=9851916;  
 RA none;  
 RT "Genome sequence of the nematode C.elegans: A platform for  
 RT investigating biology.";  
 RL Science 282:2012-2018 (1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RC Percy C.M.;  
 RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.  
 EMBL: Z93373; CAB07551.1; -.

DR PIR; T18800; T18800.  
DR WormRep; C01B9.1; CE17359.  
DR InterPro; IPR002900; DUF38.  
DR Pfam; PF01837; FTH, 4.  
KW Hypothetical protein.  
SQ SEQUENCE 1601 AA; 186279 MW; 48A2EB7F995ED995A CRC64

Query Match	50.0%	Score 7;	DB 2;	length 1601;
Best Local Similarity	100.0%	Pred. No. 1.4e+02;		
Matches 7;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY	7	ITSVLSF	13
Db	398	ITSVLSF	404

RESULT 14	
Q48626	
ID Q48626	PRELIMINARY;
	PRT; 43 AA

DT 01-NOV-1996 (TREMBLREL. 01, Created)  
DT 01-NOV-1996 (TREMBLREL. 01, Last sequence update)  
DT 01-JUN-2003 (TREMBLREL. 24, Last annotation update)  
DE ORRB of IS3-like IS.  
OS Leuconostoc lactis.  
OG Plasmid pNZ63.  
OC Bacteria; Firmicutes; Lactobacillales; Leuconostoc.  
OX NCBI\_TaxId=1246;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=6009;  
RX MEDLINE=96209221; PubMed=8633855;  
RA Vaughan E.E., David S., de Vos W.M.;  
RT "The lactose transporter in Leuconostoc lactis is a new member of the  
RT Lacs subfamily of galactoside-pentose-hexonide translocators.";  
RL Appl. Environ. Microbiol. 62:1574-1582 (1996).  
DR EMBL; U47655; AAC44115.1; --  
KW Plasmid.  
KW SEQUENCE 43 AA; 5279 MW; 440E9C6E7D461689 CRC64;

Query Match	42.9%	Score 6	DB 2	length 43
Best Local Similarity	100.0%	Pred. No. 72		
Matches 6	Conservative 0	Mismatches 0	Indels 0	Gaps 0

QY .	7	ITSVLS	12
Db	5	ITSVLS	10

RESULT 15	
Q8ZIC7	
ID Q8ZIC7	PRELIMINARY;
	PRT;
	56 AA

DT 01-MAR-2002 (TREMBLrel. 20, Created)  
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)  
DT 01-OCT-2004 (TREMBLrel. 28, Last annotation update)  
DE Hypothetical protein YP00578  
GN Ordered locus names=YP2898, YP00578;  
OS *Yersinia pestis*.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Yersiniaceae;  
OX NCBI\_TaxID=632;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA STRAIN=CO-92 / Biovar Orientalis;  
RC MEDLINE=21470413; PubMed=11586360; DOI=10.1038/35097083;  
RA Perhill U., Wren B.W., Thomson N.R., Tibbitt R.W., Holden M.T.G.,  
RA Prentice M.B., Sebahlia M., James K.D., Churchill C.M., Mungal K.L.,  
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdano, Harraga A.-M.,  
RA Chillingworth T., Cronin A., Davies R.K., Davis P., Dougan G.,  
RA Feltham T., Hamlin N., Holtroyd S., Jagsels K., Kariyasev A.V.,  
RA Leather S., Moule S., Oyston P.C.F., Quail M.A., Rutherford K.M.,  
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrall B.G.,



RC "Genome sequence of Yersinia pestis, the causative agent of plague."  
 Nature 413:523-527(2001).

RA Zhang C., Yu Y., Zhang S., Wei H., Zhou G., Ouyang S., Luo L., Bi J.,  
 Liu M., He F.,  
 RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL, AF116685; AAF71105.1; -  
 SQ SEQUENCE 81 AA; 9410 MW; 862BF81C575F6DED CRC64;

Query Match 42.9%; Score 6; DB 2; Length 81;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 IRKSKI 7  
 DB 42 IRKSKI 47

RESULT 18  
 ID P70991 PRELIMINARY; PRT; 81 AA.  
 AC P70991;  
 DT 01-FEB-1997 (TREMBLrel. 02, Created)  
 DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
 DE Hypothetical protein (Fragment).  
 OS Bacillus subtilis.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 CX NCBI\_TaxID=1423;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=168;  
 RX MEDLINE=87080286; PubMed=3098560;  
 RA Parsot C.;  
 RT "Evolution of biosynthetic pathways: a common ancestor for threonine  
 synthase, threonine dehydratase and D-serine dehydratase."  
 RL EMBL J. 5:3013-3019(1986).  
 DR EMBL, X04603; CAA28269.1; -  
 DR GO: GO:0016597; F:amino acid binding; IEA.  
 DR GO: GO:0008152; P:metabolism; IEA.  
 DR InterPro; IPR002912; ACT.  
 DR Pfam; PF01842; ACT; 1.  
 KW Hypothetical protein.  
 FT NON TER  
 SQ SEQUENCE 81 AA; 9111 MW; FA60AB27A08FB3F0 CRC64;

Query Match 42.9%; Score 6; DB 2; Length 81;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 SKITSV 10  
 DB 11 SKITSV 16

RESULT 19  
 ID O8KVS0 PRELIMINARY; PRT; 84 AA.  
 AC O8KVS0;  
 DT 01-OCT-2002 (TREMBLrel. 22, Created)  
 DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)  
 DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)  
 DE Putative early growth response protein.  
 GN Name=grp;  
 OS Renibacterium salmoninarum.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Micrococcales; Micrococcaceae; Renibacterium.  
 CX NCBI\_TaxID=1646;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=NCIMB 1114;  
 RX MEDLINE=22043279; PubMed=12047757;  
 RA Grayson T.H., Cooper L.F., Wratheini A.B., Roper J., Evenden A.J.,  
 RA Gilpin M.L.;  
 RT "Host responses to Renibacterium salmoninarum and specific components

RC TISSUB=Liver;  
 RA Zhang C., Yu Y., Zhang S., Wei H., Zhou G., Ouyang S., Luo L., Bi J.,  
 Liu M., He F.,  
 RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL, AF116685; AAF71105.1; -  
 SQ SEQUENCE 81 AA; 9410 MW; 862BF81C575F6DED CRC64;

Query Match 42.9%; Score 6; DB 2; Length 81;  
 Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 IRKSKI 7  
 DB 42 IRKSKI 47

RESULT 16  
 ID AAS63080 PRELIMINARY; PRT; 56 AA.  
 AC AAS63080;  
 DT 24-MAR-2004 (TREMBLrel. 27, Created)  
 DT 24-MAR-2004 (TREMBLrel. 27, Last sequence update)  
 DT 04-MAY-2004 (TREMBLrel. 27, Last annotation update)  
 DE Hypothetical protein.  
 GN YP2898.  
 OS Yersinia pestis.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 CC Enterobacteriaceae; Yersinia.  
 CX NCBI\_TaxID=632;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=91001 / Biovar Mediaevalis;  
 RA Song Y., Tong Z., Wang L., Han Y., Zhang J., Pei D., Wang J., Zhou D.,  
 Han Y., Pang X., Zhai J., Chen F., Qin H., Wang J., Li S., Guo Z.,  
 RA Ye C., Du Z., Lin W., Wang J., Yu J., Yang H., Wang J., Huang P.,  
 RA Yang R.;  
 RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL, AE017138; AAS63080.1; -  
 KW Hypothetical protein.  
 SQ SEQUENCE 56 AA; 6444 MW; C75FBF5B1F77AC90 CRC64;

Query Match 42.9%; Score 6; DB 2; Length 56;  
 Best Local Similarity 100.0%; Pred. No. 90;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 ITSVLS 12  
 DB 48 ITSVLS 53

RESULT 17  
 ID O9P1F0 PRELIMINARY; PRT; 81 AA.  
 AC O9P1F0;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)  
 DE PRO2049.  
 OS Homo sapiens (Human).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 CX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RT of the pathogen reveal the mechanisms of immune suppression and  
 RT activation.";  
 RL Immunology 106:273-283(2002).  
 DR EMBL; AF428072; AAM47191.1;  
 SQ SEQUENCE 84 AA; 8506 MW; 55D41CBABCE311F1 CRC64;  
 Query Match  
 Best Local Similarity 42.9%; Score 6; DB 2; Length 84;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 9 SVLSFC 14  
 63 SVLSFC 68  
 RESULT 20  
 08KHV2 PRELIMINARY; PRT; 101 AA.  
 AC 08KHV2;  
 DT 01-OCT-2002 (TREMBlrel. 22, Created)  
 DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)  
 DE 05-JUL-2004 (TREMBlrel. 27, Last annotation update)  
 DE Putative early growth response protein.  
 GN Name=grip;  
 OS Renibacterium salmoninarum.  
 OC Bacteria; Actinobacteriia; Actinobacteridae; Actinomycetales;  
 OC Micrococcineae; Micrococcaceae; Renibacterium.  
 CX NCBI\_TaxID=1646;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC33209;  
 RX MEDLINE=20087576; PubMed=10618262;  
 RA Grayson T.H., Atlenzar F.A., Alexander S.M., Cooper L.F., Gilpin M.L.;  
 RT "Molecular diversity of Renibacterium salmoninarum isolates determined  
 RT by randomly amplified polymorphic DNA analysis.";  
 RL Appl. Environ. Microbiol. 66:435-438(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC33209;  
 RX MEDLINE=22043279; PubMed=12047757;  
 RA Grayson T.H., Cooper L.F., Wrathmell A.B., Roper J., Evenden A.J.;  
 RT "Host responses to Renibacterium salmoninarum and specific components  
 RT of the pathogen reveal the mechanisms of immune suppression and  
 RT activation.";  
 RL Immunology 106:273-283(2002).  
 DR EMBL; AF178994; AAM45390.1;  
 DR EMBL; AF428071; AAM47188.1;  
 SQ SEQUENCE 101 AA; 10020 MW; 4D92137221251FCT CRC64;  
 Query Match  
 Best Local Similarity 42.9%; Score 6; DB 2; Length 101;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 9 SVLSFC 14  
 80 SVLSFC 85  
 RESULT 21  
 07OT66 PRELIMINARY; PRT; 105 AA.  
 AC 07OT66;  
 DT 01-MAR-2004 (TREMBlrel. 26, Created)  
 DT 01-MAR-2004 (TREMBlrel. 26, Last sequence update)  
 DE 01-MAR-2004 (TREMBlrel. 26, Last annotation update)  
 DE GEP 13 5047 4730.  
 OS Giardia lamblia ATCC 50803.  
 OC Eukaryota; Diplomonadida; Hexamitidae; Giardiinae; Giardia.  
 CX NCBI\_TaxID=184922;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MB C6;  
 RT Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.

RA Morrison H.G., McArthur A.G., Adam R.D., Aley S.B., Gillin F.D.,  
 RA Olsen G.U., Sogin M.L.;  
 RT "Draft sequence of the Giardia lamblia genome.";  
 RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.  
 CC -1- CAUTION: The sequence shown here is derived from an  
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
 CC preliminary data.  
 DR EMBL; AACB01000110; EAA38197.1;  
 SQ SEQUENCE 105 AA; 11562 MW; F2D35D3E9564AAB8 CRC64;  
 Query Match  
 Best Local Similarity 42.9%; Score 6; DB 2; Length 105;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Db 4 KSKITS 9  
 41 KSKITS 46  
 RESULT 22  
 0918L5 PRELIMINARY; PRT; 126 AA.  
 AC 0918L5;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DE 01-OCT-2004 (TREMBlrel. 28, Last annotation update)  
 DE Liver-basis fatty acid binding protein (Hypothetical protein).  
 GN Name=fabp10;  
 OS Brachydanio rerio (Zebrafish) (Danio rerio).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;  
 OC Cyprinidae; Danio.  
 CX NCBI\_TaxID=7955;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=20461782; PubMed=11004494;  
 RA Denovan-Wright E.M., Pierce M., Sharma M.K., Wright J.M.;  
 RT "cDNA sequence and tissue-specific expression of a basic liver-type  
 RT fatty acid binding protein in adult zebrafish (Danio rerio)."  
 RL Biochim. Biophys. Acta 1492:227-232(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Denovan-Wright E.M., Pierce M., Sharma M.K., Wright J.M.;  
 RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Liver;  
 RX MEDLINE=22386257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.U., Ueidi T.B., Toshlyuk S., Carninci P., Prange C.J.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson P.D., Miliady S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahy J., Helton E., Ketterman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
 RA Krzywinski M.I., Skalska U., Smalls D.E., Scherch A., Schein J.E.,  
 RA Jones S.J., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Liver;  
 RA Strausberg R.;  
 RT Submitted (JUL-2004) to the EMBL/GenBank/DBJ databases.

```

CC -1- SIMILARITY: Belongs to the fatty-acid binding protein (FABP)
CC family.
DR EMBL: AF54642; AAF67743.1; -.
DR EMBL: BC076219; AAH76219.1; -.
DR HSSP: P83409; 1P8P.
DR ZFIN: ZDB-GENE-020318-1; fabp10.
DR GO: GO:0008289; F:lipid binding; IEA.
DR GO: GO:0005215; F:transporter activity; IEA.
DR GO: GO:0006810; P:transport; IEA.
DR InterPro: IPR011038; Calycin.
DR InterPro: IPR000463; Fatty_acid_BP.
DR InterPro: IPR000566; Lipocin_cytfabp.
DR Pfam: PF00061; Lipocalin_N.
DR PRINTS: PR00178; FATTYACIDBP.
DR PROSITE: PS00214; FABP.1.
DR PROSITE: PS00430; TONB_DEPENDENT_REC_1; UNKNOWN_1.
DR Hypothetical protein; Transport.
SQ SEQUENCE 126 AA; 14004 MW; F77419F1F2489814 CRC64;

Query Match 42.9%; Score 6; DB 2; Length 126;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 MIRSK 6
Db 119 MIRSK 124

RESULT 23
O7UDV4 PRELIMINARY; PRT; 131 AA.
AC O7UDV4;
DT 01-OCT-2003 (TREMBlrel. 25, Created)
DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)
DE Hypothetical protein.
GN OrderedlocusNames=RB11760;
OS Rhodospirillum rubrum.
OC Bacteria; Planctomycetes; Planctomycetacia; Planctomycetales;
OC Planctomycetaceae; Pirellula.
OX NCBI_TaxID=117;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1;
RX MEDLINE=22735913; PubMed=12835416;
RA Gloeckner F.O., Kube M., Bauer M., Teeling H., Lombardot T.,
RA Ludwig W., Gade D., Beck A., Borzym K., Heltmann K., Rabus R.,
RA Schleuter H., Amann R., Reinhardt R.;
RT "Complete genome sequence of the marine planctomycete Pirellula sp.
RT strain 1."
RL Proc. Natl. Acad. Sci. U.S.A. 100:8298-8303(2003).
RW EMBL: BX294153; CAD79303.1; -.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 131 AA; 14022 MW; 1A3D469B981B1B10 CRC64;

Query Match 42.9%; Score 6; DB 2; Length 131;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 7 ITSVLS 12
Db 38 ITSVLS 43

RESULT 24
RL15 METVA STANDARD; PRT; 143 AA.
AC P14032;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE 50S ribosomal protein L15P.

```

```

GN Name=rpl15p;
OS Methanococcus vannielii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanococcaceae; Methanococcus.
OX NCBI_TaxID=2187;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90040717; PubMed=2530355;
RA Auer J., Spicker G., Beck A.;
RT "Organization and structure of the Methanococcus transcriptional unit
RT homologous to the Escherichia coli 'spectinomycin operon'.
RT Implications for the evolutionary relationship of 70 S and 80 S
RT ribosomes."
RL J. Mol. Biol. 209:21-36(1989).
CC -1- SIMILARITY: Belongs to the L15P family of ribosomal proteins.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X16720; CAA34702.1; -.
DR PIR: S05626; R6MX15.
DR InterPro: IPR001196; Ribosomal_L15.
DR Pfam: PF00256; L15; 1.
DR PROSITE: PS00475; RIBOSOMAL_L15; 1.
KW Ribosomal protein.
SQ SEQUENCE 143 AA; 15950 MW; 694468C6B3217996 CRC64;

Query Match 42.9%; Score 6; DB 1; Length 143;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 MIRSK 6
Db 1 MIRSK 6

RESULT 25
O6LXD1 PRELIMINARY; PRT; 143 AA.
AC O6LXD1;
DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
DE L15 ribosomal protein L15.
GN Name=rpl15; OrderedlocusNames=MWP1421;
OS Methanococcus maripaludis.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanococcaceae; Methanococcus.
OX NCBI_TaxID=39152;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S2 / L1;
RA Hendrickson E.L., Kaul R., Zhou Y., Bovee D., Chapman P., Chung J.,
RA Conway de Macario E., Dodsworth J., Gillett W., Graham D.E.,
RA Haydock A.K., Kang A., Land M.L., Levy R., Lie T.J., Major T.,
RA Moore B., Porat I., Overbeek R., Palmeri A., Rouse G.,
RA Saenphimachak C., Soll D., Whitman W.B., Larimer F.W., Olson M.V.,
RA Leigh J.A.;
RT "Complete genome sequence of the mesophilic hydrogenotrophic
RT methanogen Methanococcus maripaludis."
RT Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Belongs to the L15P family of ribosomal proteins.
DR EMBL: BX57222; CAF30977.1; -.
DR GO: GO:0003735; F:structural constituent of ribosome; IEA.
DR InterPro: IPR001196; Ribosomal_L15.
DR Pfam: PF00256; L15; 1.
DR PROSITE: PS00475; RIBOSOMAL_L15; 1.
KW Complete proteome; Ribonucleoprotein; Ribosomal protein.

```

SQ SEQUENCE 143 AA; 15951 MW; AACFF8148932FB90 CRC64;

Query Match 42.9%; Score 6; DB 2; Length 143;  
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MIRKSK 6  
 |||||  
 Db 1 MIRKSK 6

RESULT 26  
 CAF30977 PRELIMINARY; PRT; 143 AA.

ID Q7VC75  
 AC CAF30977;  
 DT 03-MAR-2004 (TREMBlrel. 27, Last sequence update)  
 DT 03-MAR-2004 (TREMBlrel. 27, Last sequence update)  
 DE L5U ribosomal protein L15.  
 GN RPL0 OR MPM1421.  
 OS Methanococcus maripaludis.  
 OC Archaea; Euryarchaeota; Methanococci; Methanococcales;  
 OC Methanococcaceae; Methanococcus.  
 CX NCBI\_TaxID=39152;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S2 / IL;  
 RA Hendrickson E.L., Kaul R., Zhou Y., Boyce D., Chapman P., Chung J.,  
 RA Conway de Macario E., Dodsworth J., Gillette W., Graham D.E.,  
 RA Haydock A.K., Kang A., Land M.L., Levy R., Lie T.J., Major T.,  
 RA Moore B., Porat I., Overbeek R., Palmelrt A., Rouse G.,  
 RA Saenphimachak C., Soell D., Whitman W.B., Laxtner F.W., Olson M.V.,  
 RA Leigh J.A.;  
 RT "Complete genome sequence of the mesophilic hydrogentrophic  
 RT methanogen Methanococcus maripaludis";  
 RL Submitted (FEB-2004) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; BX957222; CAF30977.1; -.  
 KW Ribosomal protein.  
 SQ SEQUENCE 143 AA; 15951 MW; AACFF8148932FB90 CRC64;

Query Match 42.9%; Score 6; DB 2; Length 143;  
 Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MIRKSK 6  
 |||||  
 Db 1 MIRKSK 6

RESULT 27  
 Q7VC75 PRELIMINARY; PRT; 153 AA.

ID Q7VC75  
 AC Q7VC75;  
 DT 01-OCT-2003 (TREMBlrel. 25, Created)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)  
 DE Predicted endonuclease.  
 GN OrderedLocustNames=Pro0867;  
 OS Prochlorococcus marinus.  
 OC Bacteria; Cyanobacteria; Prochlorophytes; Prochlorococcaceae;  
 OC Prochlorococcus.  
 CX NCBI\_TaxID=1219;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SARG / CCM 1375 / SS120;  
 RX MEDLINE=22810154; PubMed=12917486;  
 RA Diferene A., Salanoubat M., Partensky F., Artiguenave F., Axmann I.M.,  
 RA Barbe V., Duprat S., Galperin M.Y., Koonin E.V., Le Gall F.,  
 RA Sekatova K.S., Ostrowski M., Ozlas S., Robert C., Rogozin I.B.,  
 RA Scanlan D.J., Tandeau de Marsac N., Weissenbach J., Wincker P.,  
 RA Wolf Y.I., Hess W.R.;  
 RT "genome sequence of the cyanobacterium Prochlorococcus marinus SS120,  
 a nearly minimal oxyphototrophic genome.";

RL Proc. Natl. Acad. Sci. U.S.A. 100:10020-10025(2003).

DR EMBL; AE017163; AAP9911.1; -.  
 DR GO: GO:0004519; F:endonuclease activity; IEA.  
 DR GO: GO:0016788; F:hydrolase activity, acting on ester bonds; IEA.  
 DR GO: GO:0006139; P:nucleobase, nucleoside, nucleotide and nucl. . .; IEA.  
 DR InterPro; IPR005227; Cons\_hypoth250.  
 DR InterPro; IPR006641; YGFPC.  
 DR Pfam; PF03652; UPF0081.1.  
 DR SMART; SM00732; YGFPC; 1.  
 DR TIGRFAMs; TIGR00250; cons\_hypoth250; 1.  
 KW Complete proteome; Endonuclease.  
 SQ SEQUENCE 153 AA; 17000 MW; 19D3E9F78FA3CB69 CRC64;

Query Match 42.9%; Score 6; DB 2; Length 153;  
 Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 TSVLSF 13  
 |||||  
 Db 7 TSVLSF 12

RESULT 28  
 Q7RKH6 PRELIMINARY; PRT; 157 AA.

ID Q7RKH6  
 AC Q7RKH6;  
 DT 01-MAR-2004 (TREMBlrel. 26, Created)  
 DT 01-MAR-2004 (TREMBlrel. 26, Last sequence update)  
 DE Hypothetical protein.  
 GN Name=PY02925;  
 OS Plasmodium yoelii yoelii.  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 CX NCBI\_TaxID=73239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=17XNL;  
 RX PubMed=12368865;  
 RA Carlton J.M., Angiolini S.V., Suh B.B., Kooij T.W., Perce M.,  
 RA Silva J.C., Emolaeva M.D., Allen J.E., Sengenut J.D., Koo H.L.,  
 RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,  
 RA Shallow S.J., van Aken S.E., Riedmiller S.B., Feldblyum T.V.,  
 RA Cho J.K., Quackenbush J., Sedegah M., Shoib A., Cummings L.M.,  
 RA Florens L., Yates F.R. III, Raine J.D., Sinden R.E., Harris M.A.,  
 RA Cunningham D.A., Preiser P.R., Bergman L.W., Valdivia A.B.,  
 RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,  
 RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,  
 RA Carucci D.J.;  
 RT "genome sequence and comparative analysis of the model rodent malaria  
 RT parasite Plasmodium yoelii yoelii";  
 RL Nature 419:512-519(2002).  
 CC -!- CAUTION: The sequence shown here is derived from an  
 CC EMBL/Genbank/DBJ whole genome shotgun (WGS) entry which is  
 CC preliminary data.  
 DR EMBL; AABL0100823; EAA22443.1; -.  
 DR GO: GO:0009331; C:glycerol-3-phosphate dehydrogenase complex; IEA.  
 DR GO: GO:0004367; F:glycerol-3-phosphate dehydrogenase (NAD+) a. . .; IEA.  
 DR GO: GO:0016614; F:oxidoreductase activity, acting on CH-OH gr. . .; IEA.  
 DR GO: GO:0005975; P:carbohydrate metabolism; IEA.  
 DR GO: GO:0006072; P:glycerol-3-phosphate metabolism; IEA.  
 DR InterPro; IPR006109; NAD\_Gly3P\_C.  
 DR InterPro; IPR006168; NAD\_Gly3P\_dh.  
 DR Pfam; PF07479; NAD\_Gly3P\_dh\_C; 1.  
 DR PRINTS; PR00077; GPDHGRNASE.  
 DR ProDom; PD001278; NAD\_Gly3P\_C; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 157 AA; 17683 MW; 14E354CF507B4DD CRC64;

Query Match 42.9%; Score 6; DB 2; Length 157;  
 Best Local Similarity 100.0%; Pred. No. 2.1e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 TSVLS 12

Db 141 ITSvLS 146

## RESULT 29

Q737A5 PRELIMINARY; PRT; 167 AA.  
AC Q737A5;  
DT 05-JUL-2004 (TrEMBLrel. 27, Created)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
DE Hypothetical protein.  
GN OrderedLocustNames=BCE2745;  
OS *Bacillus cereus* (strain ATCC 10987).  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=222523;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX PubMed=14960714;  
RA Raeko D.A., Ravel J., Oekstad O.A., Helgason E., Cer R.Z., Jiang L.,  
Shores K.A., Fouts D.E., Tourasse N.J., Angiuoli S.V., Kolonay J.F.,  
Neilson W.C., Kolstoe A.-B., Fraser C.M., Read T.D.;  
RT "The genome sequence of *Bacillus cereus* ATCC 10987 reveals metabolic  
adaptations and a large plasmid related to *Bacillus anthracis* pXO1.";  
RL Nucleic Acids Res. 32:977-988(2004).  
DR EMBL; AE017273; AAS41657.1; -.  
DR TIGR; BCE2745; -.  
DR InterPro; IPR011235; DUP UCP032285.  
DR PIRSF; PIRSF032285; UCP032285; 1.  
KW Complete proteome; Hypothetical protein.  
SQ SEQUENCE 167 AA; 19611 MW; 73539A73CFDD18B7 CRC64;

Query March 42.9%; Score 6; DB 2; Length 167;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 RKSkit 8  
Db 50 RKSkit 55

RESULT 30  
AAS41657 PRELIMINARY; PRT; 167 AA.  
AC AAS41657;  
DT 02-MAR-2004 (TrEMBLrel. 27, Created)  
DT 02-MAR-2004 (TrEMBLrel. 27, Last sequence update)  
DT 04-MAY-2004 (TrEMBLrel. 27, Last annotation update)  
DE Hypothetical protein.  
GN BCE2745.  
OS *Bacillus cereus* (strain ATCC 10987).  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=222523;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX PubMed=14960714;  
RA Raeko D.A., Ravel J., Oekstad O.A., Helgason E., Cer R.Z., Jiang L.,  
Shores K.A., Fouts D.E., Tourasse N.J., Angiuoli S.V., Kolonay J.F.,  
Neilson W.C., Kolstoe A.-B., Fraser C.M., Read T.D.;  
RT "The genome sequence of *Bacillus cereus* ATCC 10987 reveals metabolic  
adaptations and a large plasmid related to *Bacillus anthracis* pXO1.";  
RL Nucleic Acids Res. 32:977-988(2004).  
DR EMBL; AE017273; AAS41657.1; -.  
DR TIGR; BCE2745; -.  
KW Hypothetical protein.  
SQ SEQUENCE 167 AA; 19611 MW; 73539A73CFDD18B7 CRC64;

Query March 42.9%; Score 6; DB 2; Length 167;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 RKSkit 8  
Db 50 RKSkit 55

Db 50 RKSkit 55

## RESULT 31

Q70GJ2 PRELIMINARY; PRT; 169 AA.  
AC Q70GJ2;  
DT 05-JUL-2004 (TrEMBLrel. 27, Created)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
DE Nucleotide binding site leucine-rich repeat disease resistance protein  
(Fragment).  
OS *Pyrus communis* (Pear).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC eurossids I; Rosales; Rosaceae; Maloideae; Pyrus.  
OX NCBI\_TaxID=23211;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Afunian M.R.;  
RL Submitted (Aug-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ581777; CAB46476.1; -.  
FT NON TER 1  
FT NON TER 1  
SQ SEQUENCE 169 AA; 19132 MW; 38342B7F867B351B CRC64;

Query March 42.9%; Score 6; DB 2; Length 169;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 SKITSV 10  
Db 51 SKITSV 56

## RESULT 32

Q70GJ3 PRELIMINARY; PRT; 169 AA.  
AC Q70GJ3;  
DT 05-JUL-2004 (TrEMBLrel. 27, Created)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
DE Nucleotide binding site leucine-rich repeat disease resistance protein  
(Fragment).  
OS *Pyrus communis* (Pear).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC eurossids I; Rosales; Rosaceae; Maloideae; Pyrus.  
OX NCBI\_TaxID=23211;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Afunian M.R.;  
RL Submitted (Aug-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ581776; CAB46475.1; -.  
FT NON TER 1  
FT NON TER 1  
SQ SEQUENCE 169 AA; 19141 MW; 33971B1C03BF9620 CRC64;

Query March 42.9%; Score 6; DB 2; Length 169;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 SKITSV 10  
Db 51 SKITSV 56

## RESULT 33

CAB46475 PRELIMINARY; PRT; 169 AA.  
AC CAB46475;  
DT 02-MAR-2004 (TrEMBLrel. 27, Created)  
DT 02-MAR-2004 (TrEMBLrel. 27, Last sequence update)

DT 02-MAR-2004 (TREMBlrel. 27, Last annotation update)  
DE Nucleotide binding site leucine-rich repeat disease resistance protein  
(Fragment).  
OS Pyrus communis (Pear).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC eurossids I; Rosales; Rosaceae; Maloideae; Pyrus.  
OX NCBI\_TaxID=23211;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Seckel;  
RA Afunian M.R.;  
RT "Molecular approaches to improving disease resistance in apple and  
RT pear";  
RL Submitted (AUG-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL, AJ581776; CAB6475.1; -.  
FT NON\_TER 1 1  
SQ SEQUENCE 169 AA; 19141 MW; 33971B1C03BF620 CRC64;

Query Match 42.9%; Score 6; DB 2; Length 169;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 SKITSV 10  
DB 51 SKITSV 56

## RESULT 34

CAB6476 PRELIMINARY; PRT; 169 AA.

AC CAB6476;  
DT 02-MAR-2004 (TREMBlrel. 27, Created)  
DT 02-MAR-2004 (TREMBlrel. 27, Last sequence update)  
DT 02-MAR-2004 (TREMBlrel. 27, Last annotation update)  
DE Nucleotide binding site leucine-rich repeat disease resistance protein  
(Fragment).  
OS Pyrus communis (Pear).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC eurossids I; Rosales; Rosaceae; Maloideae; Pyrus.  
OX NCBI\_TaxID=23211;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Seckel;  
RA Afunian M.R.;  
RT "Molecular approaches to improving disease resistance in apple and  
RT pear";  
RL Submitted (AUG-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL, AJ581777; CAB6476.1; -.  
FT NON\_TER 1 1  
SQ SEQUENCE 169 AA; 19132 MW; 38342B7F867B351B CRC64;

Query Match 42.9%; Score 6; DB 2; Length 169;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 SKITSV 10  
DB 51 SKITSV 56

## RESULT 35

Q70G16 PRELIMINARY; PRT; 170 AA.

AC Q70G16;  
DT 05-JUL-2004 (TREMBlrel. 27, Created)  
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)  
DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)  
DE Nucleotide binding site leucine-rich repeat disease resistance protein  
(Fragment).

OS Pyrus communis (Pear).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC eurossids I; Rosales; Rosaceae; Maloideae; Pyrus.  
OX NCBI\_TaxID=23211;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Seckel;  
RA Afunian M.R.;  
RT "Molecular approaches to improving disease resistance in apple and  
RT pear";  
RL Submitted (AUG-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL, AJ581783; CAB6655.1; -.  
FT NON\_TER 1 1  
SQ SEQUENCE 170 AA; 19334 MW; 73547B4FE1B89C2F CRC64;

Query Match 42.9%; Score 6; DB 2; Length 170;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 SKITSV 10  
DB 52 SKITSV 57

## RESULT 36

CAB6655 PRELIMINARY; PRT; 170 AA.

AC CAB6655;  
DT 02-MAR-2004 (TREMBlrel. 27, Created)  
DT 02-MAR-2004 (TREMBlrel. 27, Last sequence update)  
DT 02-MAR-2004 (TREMBlrel. 27, Last annotation update)  
DE Nucleotide binding site leucine-rich repeat disease resistance protein  
(Fragment).  
OS Pyrus communis (Pear).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC eurossids I; Rosales; Rosaceae; Maloideae; Pyrus.  
OX NCBI\_TaxID=23211;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Seckel;  
RA Afunian M.R.;  
RT "Molecular approaches to improving disease resistance in apple and  
RT pear";  
RL Submitted (AUG-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL, AJ581783; CAB6655.1; -.  
FT NON\_TER 1 1  
SQ SEQUENCE 170 AA; 19334 MW; 73547B4FE1B89C2F CRC64;

Query Match 42.9%; Score 6; DB 2; Length 170;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 SKITSV 10  
DB 52 SKITSV 57

## RESULT 37

Q9L362 PRELIMINARY; PRT; 173 AA.

ID Q9L362;  
DT 01-OCT-2000 (TREMBlrel. 15, Created)  
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)  
DE Aerobic ribonucleotide reductase class Ib (EC 1.17.4.1)  
(Fragment).  
GN Name=nrde;  
OS Yersinia pseudotuberculosis.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Yersinia.  
OX NCBI\_TaxID=633;  
RN [1]

RP SEQUENCE FROM N.A.  
 RA Torrens E., Jordan A., Gilbert I., Karlsson M.;  
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Margaretta K.;  
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
 RN EMBL: AJ286852; CAB87141.1;  
 DR GO: GO:0005971; C:ribonucleoside-diphosphate reductase complex; IEA.  
 DR GO: GO:0016491; F:oxidoreductase activity; IEA.  
 DR GO: GO:0004748; F:ribonucleoside-diphosphate reductase activity; IEA.  
 DR GO: GO:0006260; P:DNA replication; IEA.  
 DR InterPro: IPR00788; Ribonucleo red.  
 DR Pfam: PF02867; Ribonuc red. lgc; 1.  
 DR PRINTS: PR01183; RIBORDTASEM1.  
 KW Oxidoreductase.  
 FT NON\_TER 1 1  
 FT NON\_TER 173 173  
 SQ SEQUENCE 173 AA; 19697 MW; 33C10E33B458311A CRC64;

Query Match 42.9%; Score 6; DB 2; Length 173;  
 Best Local Similarity 100.0%; Pred. No. 2.3e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 2 IRKSKI 7  
 |||||  
 Db 43 IRKSKI 48

## RESULT 38

09L364 PRELIMINARY; PRT; 173 AA.  
 AC 09L364;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)  
 DE Aerobic ribonucleotide reductase class Ib (EC 1.17.4.1)  
 DE (Fragment).  
 GN Name: rrd.  
 OS Versinia pectis.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Yersinia.  
 OX NCBI\_TaxID=632;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Torrens E., Jordan A., Gilbert I., Karlsson M.;  
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Margaretta K.;  
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
 RN EMBL: AJ286850; CAB87139.1;  
 DR GO: GO:0005971; C:ribonucleoside-diphosphate reductase complex; IEA.  
 DR GO: GO:0016491; F:oxidoreductase activity; IEA.  
 DR GO: GO:0004748; F:ribonucleoside-diphosphate reductase activity; IEA.  
 DR GO: GO:0006260; P:DNA replication; IEA.  
 DR InterPro: IPR00788; Ribonucleo red.  
 DR Pfam: PF02867; Ribonuc red. lgc; 1.  
 DR PRINTS: PR01183; RIBORDTASEM1.  
 KW Oxidoreductase.  
 FT NON\_TER 1 1  
 FT NON\_TER 173 173  
 SQ SEQUENCE 173 AA; 19697 MW; 33C10E33B458311A CRC64;

Query Match 42.9%; Score 6; DB 2; Length 173;  
 Best Local Similarity 100.0%; Pred. No. 2.3e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 2 IRKSKI 7  
 |||||  
 Db 43 IRKSKI 48

RESULT 39  
 06HI08 PRELIMINARY; PRT; 174 AA.  
 AC 06HI08;  
 DT 05-JUL-2004 (TREMBlrel. 27, Created)  
 DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)  
 DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)  
 DE Hypothetical protein.  
 GN ORFNames=BT9727.2492;  
 OS Bacillus thuringiensis serovar konkukian str. 97-27.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus;  
 OC Bacillus thuringiensis serovar konkukian.  
 OX NCBI\_TaxID=281309;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=97-27;  
 RA Bretlin T.S., Bruce D., Challacombe J.F., Gilna P., Han C., Hill K.,  
 RA Hitchcock P., Jackson P., Keim P., Longmire J., Lucas S., Okinaka R.,  
 RA Richardson P., Rubin E., Tice H.;  
 RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AB017355; AAT61854.1;  
 DR InterPro: IPR011235; DUF\_UCP032285.  
 DR PIRSF: PIRSF032285; UCP032285; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 174 AA; 20427 MW; 43AFABBF23FCB2 CRC64;

Query Match 42.9%; Score 6; DB 2; Length 174;  
 Best Local Similarity 100.0%; Pred. No. 2.3e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 3 RKSKIT 8  
 |||||  
 Db 57 RKSKIT 62

## RESULT 40

08IPS9 PRELIMINARY; PRT; 174 AA.  
 AC 08IPS9;  
 DT 01-JUN-2003 (TREMBlrel. 24, Created)  
 DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)  
 DT 01-OCT-2004 (TREMBlrel. 28, Last annotation update)  
 DE Hypothetical protein.  
 GN OrderedLocusNames=BA2718, BAS2532; ORFNames=GMAA2718;  
 OS Bacillus anthracis.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
 OX NCBI\_TaxID=1392;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Ames / isolate Porton;  
 RX MEDLINE=22608414; PubMed=12721629; DOI=10.1038/nature01586;  
 RA Read T.D., Peterson S.N., Tourasse N.J., Baillie L.W., Paulsen I.T.,  
 RA Nelson K.E., Tettelin H., Fouts D.E., Eisen J.A., Gill S.R.,  
 RA Holtzapple E.K., Ostad O.A., Helgason E., Ristone J., Wu M.,  
 RA Kolonay J.F., Beaman M.J., Dodson R.J., Brinkac L.M., Gilm M.L.,  
 RA Deboy R.T., Madpu R., Daugherty S.C., Durkin A.S., Haft D.H.,  
 RA Nelson W.C., Peterson J.D., Pop M., Khouri H.M., Radune D.,  
 RA Benton J.L., Mahamoud Y., Jiang L., Hance I.R., Weidman J.F.,  
 RA Berry K.J., Plaut R.D., Wolf A.M., Watkins J.L., Nieman W.C.,  
 RA Hazen A., Cline R.T., Redmond C., Thwaitte J.E., White O.,  
 RA Salzberg S.L., Thomson B., Friedlander A.M., Koehler T.M.,  
 RA Hanna P.C., Kolstoe A.B., Fraser C.M.;  
 RT "The genome sequence of Bacillus anthracis Ames and comparison to  
 RT closely related bacteria."  
 RL Nature 423:81-86 (2003).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Ames / isolate 0581;  
 RA Ravel J., Rasko D.A., Shumway M.F., Jiang L., Cer R.Z., Federova N.B.,  
 RA Wilson M., Stanley S., Decker S., Read T.D., Salzberg S.L.,  
 RA Fraser C.M.;  
 RT "Bacillus anthracis comparative genomics."  
 RT Submitted (MAY-2004) to the EMBL/GenBank/DBJ databases.

```

RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Stene; Bruce D., Challacombe J.F., Gilna P., Han C., Hill K.,
RA Brettin T.S., Jackson P., Keim P., Longmire J., Lucas S., Okinaka R.,
RA Hitchcock P., Rubin E., Tice H.;
RA Richardson P., Rubin E., Tice H.;
RL Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE017032; AAP26558.1; -
DR EMBL; AE017334; AAT1833.1; -
DR EMBL; AE017225; AAT54843.1; -
DR TIGR; BA2718; -
DR InterPro; IPR01235; DUF UCP032285.
DR PIRSF; PIRSF032285; UCP032285; 1.
KW Hypothetical protein.
SQ SEQUENCE 174 AA; 20425 MW; BA90063C0DC85153 CRC64;

Query Match 42.9%; Score 6; DB 2; Length 174;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 RSKIT 8
Db 57 RSKIT 62

```

Search completed: December 30, 2004, 15:37:17  
 Job time : 199 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

# OM protein - protein search, using sw model

Run on: December 30, 2004, 15:16:09 : Search time 38 Seconds  
(without alignments)  
35.448 Million cell updates/sec

Title: US-10-719-385-20  
Perfect score: 14  
Sequence: 1 MIRSKITSVLSFC 14

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 283416 seqs, 96216763 residues

Word size : 0

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 150 summaries

Database :

1: PIR.79:.\*  
2: PIR1:.\*  
3: PIR3:.\*  
4: PIR4:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	7	50.0	1601	2 T18800	hypothetical prote
2	6	42.9	56	2 AH0071	hypothetical prote
3	6	42.9	75	2 T4139	hypothetical prote
4	6	42.9	143	1 R6MX15	ribosomal protein
5	6	42.9	159	2 T15627	hypothetical prote
6	6	42.9	184	1 TPRBIV	troponin I, slow 8
7	6	42.9	187	2 B44786	troponin I, slow 8
8	6	42.9	200	2 C49600	coat protein 22k -
9	6	42.9	255	2 T25853	hypothetical prote
10	6	42.9	316	2 D89830	hypothetical prote
11	6	42.9	324	2 S70390	DMC1/LIM5 homolog
12	6	42.9	326	2 T51811	protein farnesyltr
13	6	42.9	327	2 T00797	hypothetical prote
14	6	42.9	340	2 T49006	farnesyltransferas
15	6	42.9	343	2 A75597	mannosyltransferas
16	6	42.9	404	2 D95233	aminotransferase,
17	6	42.9	404	2 P98097	aspartate transami
18	6	42.9	429	2 P86240	hypothetical prote
19	6	42.9	433	1 DESCH8	homoserine dehydro
20	6	42.9	433	2 A83728	adenylosuccinate 1
21	6	42.9	469	2 D84949	NADH2 dehydrogenas
22	6	42.9	482	2 G86227	hypothetical prote
23	6	42.9	543	2 S45492	isp5 protein - fts
24	6	42.9	560	2 C44632	hypothetical prote
25	6	42.9	566	2 S19063	hypothetical prote
26	6	42.9	572	2 AF0211	probable ABC trans
27	6	42.9	580	2 T50059	sexual differentiation
28	6	42.9	587	2 S58319	hypothetical prote
29	6	42.9	644	2 D85359	hypothetical prote

30	6	42.9	651	2 G64068	DNA topoisomerase
31	6	42.9	652	2 H86221	hypothetical prote
32	6	42.9	693	2 AC0323	ribonucleoside-dip
33	6	42.9	727	2 T23585	hypothetical prote
34	6	42.9	809	2 C96552	hypothetical prote
35	6	42.9	809	2 T41645	probable spindle p
36	6	42.9	840	2 T33217	hypothetical prote
37	6	42.9	862	2 T49583	differentiation an
38	6	42.9	909	2 G69599	aconitate hydratase
39	6	42.9	1205	2 T27053	hypothetical prote
40	6	42.9	1242	2 S78061	DNA-directed RNA p
41	6	42.9	1377	2 D90538	hypothetical prote
42	6	42.9	1385	2 H88569	protein K03H1.5 [i
43	6	42.9	1409	2 S41028	hypothetical prote
44	6	42.9	1603	2 D89407	protein R10B8.6 [i
45	6	42.9	1696	2 T24146	hypothetical prote
46	5	35.7	53	2 S41957	naringenin-chalcon
47	5	35.7	56	2 S45027	sox1 protein - Her
48	5	35.7	57	1 W9BPC7	gene 19.3 protein
49	5	35.7	59	2 G91095	hypothetical prote
50	5	35.7	59	2 C85941	hypothetical prote
51	5	35.7	60	2 D97849	hypothetical prote
52	5	35.7	70	2 T17923	hypothetical prote
53	5	35.7	72	1 OHEC1	heat-stable entero
54	5	35.7	72	1 OHEC4	heat-stable entero
55	5	35.7	72	1 OHEC4	heat-stable entero
56	5	35.7	80	2 OHECIB	heat-stable entero
57	5	35.7	80	2 D69745	hypothetical prote
58	5	35.7	80	2 A85537	unknown [imported]
59	5	35.7	83	2 H71068	hypothetical prote
60	5	35.7	87	2 AF0910	conserved hypochet
61	5	35.7	92	2 E95163	conserved hypochet
62	5	35.7	92	2 D98029	conserved hypochet
63	5	35.7	96	2 F64319	hypothetical prote
64	5	35.7	98	2 B97821	hypothetical prote
65	5	35.7	98	2 A97632	hypothetical prote
66	5	35.7	99	2 AD3031	hypothetical prote
67	5	35.7	100	1 S26080	ribosomal protein
68	5	35.7	103	2 C75149	hypothetical prote
69	5	35.7	104	2 H65115	hypothetical prote
70	5	35.7	104	2 B85968	hypothetical prote
71	5	35.7	104	2 G91142	hypothetical prote
72	5	35.7	105	2 F86774	hypothetical prote
73	5	35.7	106	2 A69027	ribosomal protein
74	5	35.7	107	2 A49442	Ig heavy chain V r
75	5	35.7	107	2 S61135	hypothetical prote
76	5	35.7	107	2 T51124	lycopen cyclase [
77	5	35.7	109	1 PVRVC	parvalbumin - thox
78	5	35.7	113	2 S26465	Ig heavy chain V r
79	5	35.7	114	2 B83798	hypothetical prote
80	5	35.7	115	2 D64710	hypothetical prote
81	5	35.7	115	2 A71809	hypothetical prote
82	5	35.7	116	2 S68195	probable membrane
83	5	35.7	117	2 C69969	hypothetical prote
84	5	35.7	118	1 B64248	hypothetical prote
85	5	35.7	122	2 S11740	Ig heavy chain pre
86	5	35.7	122	2 S69460	hypothetical prote
87	5	35.7	125	2 JU0179	heparin-binding pr
88	5	35.7	126	2 JC7571	fatty acid-binding
89	5	35.7	128	2 S52084	riboseomal protein
90	5	35.7	128	2 S33611	naringenin-chalcon
91	5	35.7	128	2 JC2120	heparin-binding pr
92	5	35.7	128	2 JC2121	heparin-binding pr
93	5	35.7	128	2 JC2119	heparin-binding pr
94	5	35.7	130	2 H86832	hypothetical prote
95	5	35.7	132	1 T43826	ribosomal protein
96	5	35.7	132	2 F84332	50S ribosomal prot
97	5	35.7	132	2 B84543	actin depolymerizi
98	5	35.7	136	2 A69461	vinulance associat
99	5	35.7	136	2 B90058	conserved hypochet
100	5	35.7	137	2 A11176	E. coli PhnB prote
101	5	35.7	142	2 B90155	hypothetical prote
102	5	35.7	142	2 QJ1865	hypothetical 15.1K

103 35.7 143 2 PT0174 Ig heavy chain Drc  
104 35.7 143 2 JN0093 hypothetical 17K p  
105 35.7 143 2 A71050 hypothetical prote  
106 35.7 145 1 U00347 capsid protein VP3  
107 35.7 145 2 G69011 hypothetical prote  
108 35.7 145 2 AD3143 conserved hypotnet  
109 35.7 145 2 H98144 hypothetical prote  
110 35.7 146 2 AE09822 probable acetyltra  
111 35.7 147 2 AB4546 50S ribosomal prot  
112 35.7 148 2 AG1645 hypothetical prote  
113 35.7 148 2 AE1594 hypothetical prote  
114 35.7 150 2 S75504 glycolytic prote  
115 35.7 152 2 T07858 glycine-rich prote  
116 35.7 157 2 G82080 hypothetical prote  
117 35.7 160 2 A70815 probable moq prote  
118 35.7 162 2 T49957 ribosomal protein  
119 35.7 166 2 G86167 protein F2187.33 l  
120 35.7 167 2 T00888 hypothetical prote  
121 35.7 169 2 AE1166 transcription regu  
122 35.7 170 2 T11964 allophycocyanin be  
123 35.7 171 2 D66444 probable heat shoc  
124 35.7 174 2 G70220 exported protein A  
125 35.7 175 2 G72683 hypothetical prote  
126 35.7 175 2 G96532 hypothetical prote  
127 35.7 175 2 T28873 hypothetical prote  
128 35.7 176 2 H70196 adenine phosphorib  
129 35.7 177 2 T25232 hypothetical prote  
130 35.7 177 2 H80299 hypothetical prote  
131 35.7 179 2 B90408 hypothetical prote  
132 35.7 180 2 B96619 protein T30316.5 l  
133 35.7 181 2 B90588 hypothetical prote  
134 35.7 181 2 AF1913 hypothetical prote  
135 35.7 181 2 B69844 lytic transglycosy  
136 35.7 182 2 T28270 ORF MSV110 hypotne  
137 35.7 182 2 H80288 conserved hypotne  
138 35.7 187 1 E70443 L-fuculose-phospha  
139 35.7 191 2 S69735 hypothetical prote  
140 35.7 197 2 JC7735 frizzled-related p  
141 35.7 202 2 F97342 transcription regu  
142 35.7 203 2 T20111 hypothetical prote  
143 35.7 203 2 A43856 major fibrillar pro  
144 35.7 206 2 B76189 hypothetical prote  
145 35.7 209 2 C97461 probable marf-fam1  
146 35.7 212 2 B89891 hypothetical prote  
147 35.7 213 2 S30181 H-transporter (a  
148 35.7 213 2 AB3685 ABC transporter (a  
149 35.7 215 2 S49326 nascent polypeptid  
150 35.7 215 2 T30827 nascent polypeptid

## ALIGNMENTS

## RESULT 1

T18800 hypothetical protein C01B9.1 - Caenorhabditis elegans

C/Species: Caenorhabditis elegans

C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004

C/Accession: T18800

R/percy C.

submitted to the EMBL Data Library, March 1997

A/Reference number: Z19024

A/Accession: T18800

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-1601 <WIL>

A/Cross-references: UNIPROT:O45231, EMBL:Z93373, P1DN:CA807551.1, GSPDB:GN00020, CESP:CD

A/Experimental source: clone C01B9

C/Genes: CESP:C01B9.1

A/Map position: 2

A/Intons: 151/3; 243/2; 478/3; 569/2; 795/3; 879/3; 895/2; 1227/3; 1270/2; 1514/3

Query Match 50.0%; Score 7; DB 2; Length 1601;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ITSVLSF 13  
DB 398 ITSVLSF 404

## RESULT 2

AH0071 hypothetical protein YP00578 [imported] - Yersinia pestis (strain CO92)

C/Species: Yersinia pestis

C/Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 09-Jul-2004

C/Accession: AH0071

R/Parkhill, J.; Wren, B.W.; Thomson, N.R.; Tilball, R.W.; Holden, M.T.G.; Prentice, M.B.;

deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;

11, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall, J.

Nature 413, 523-527, 2001

A/Title: Genome sequence of Yersinia pestis, the causative agent of plague.

A/Reference number: AB0001; MUID:21470413; PMID:11586360

A/Accession: AH0071

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-56 <KUR>

A/Cross-references: UNIPROT:Q8ZIC7; GB:AL590842; P1DN:CA89435.1; P1D:G15978671; GSPDB:G

C/Genes: YP00578

Query Match 42.9%; Score 6; DB 2; Length 56;  
Best Local Similarity 100.0%; Pred. No. 14;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ITSVLS 12  
DB 48 ITSVLS 53

## RESULT 3

T44139 hypothetical protein [imported] - Staphylococcus aureus (fragment)

C/Species: Staphylococcus aureus

C/Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 11-May-2000

C/Accession: T44139

R/Ito, T.; Katayama, Y.; Hiramoto, K.

Antimicrob. Agents Chemother. 43, 1449-1458, 1999

A/Title: Cloning and nucleotide sequence determination of the entire mec DNA of pre-meth

A/Reference number: Z22733; MUID:99278010; PMID:10348769

A/Accession: T44139

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-75 <IRO>

A/Cross-references: EMBL:DB6934; P1DN:BA82241.1

A/Experimental source: strain N315

Query Match 42.9%; Score 6; DB 2; Length 75;  
Best Local Similarity 100.0%; Pred. No. 18;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 SVLSFC 14  
DB 48 SVLSFC 53

## RESULT 4

R6X15 ribosomal protein L15 - Methanococcus vannielii

C/Species: Methanococcus vannielii

C/Date: 31-Mar-1991 #sequence\_revision 31-Mar-1991 #text\_change 09-Jul-2004

C/Accession: S05626

R/Anet, U.; Spicker, G.; Boeck, A.

U. Mol. Biol. 209, 21-36, 1989

A/Title: Organization and structure of the Methanococcus transcriptional unit homologous

## S ribosome.

A:Reference number: S05611; MUID:90040717; PMID:2530355  
 A:Accession: S05626  
 A:Molecule type: DNA  
 A:Residues: 1-143 <ABE>  
 A:Cross-references: UNIPROT:P14032; EMBL:X16720; NID:g44754; PIDN:CAA34702.1; PID:g44770  
 C:Superfamily: rat ribosomal protein L27a  
 C:Keywords: protein biosynthesis; ribosome

Query Match 42.9%; Score 6; DB 1; Length 143;  
 Best Local Similarity 100.0%; Pred. No. 30;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MIRRSK 6  
 Db 1 MIRRSK 6

## RESULT 5

T15627  
 hypothetical protein C25H3.5 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans  
 C>Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 21-Jan-2000  
 C:Accession: T15627  
 R:Johnson, D.  
 submitted to the EMBL Data Library, June 1995  
 A:Description: The sequence of C. elegans cosmid C25H3.  
 A:Reference number: Z18379  
 A:Accession: T15627  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-159 <JOH>  
 A:Cross-references: EMBL:U29535; NID:g860251; PID:AAA68784.1; CESP:C25H3.5  
 A:Experimental source: strain Bristol N2  
 C:Genetics:  
 A:Gene: CESP:C25H3.5  
 A:Insertions: 21/1  
 C:Superfamily: Caenorhabditis elegans hypothetical protein C25H3.5

Query Match 42.9%; Score 6; DB 2; Length 159;  
 Best Local Similarity 100.0%; Pred. No. 33;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MIRRSK 6  
 Db 106 MIRRSK 111

## RESULT 6

TPRBIW

troponin I, slow skeletal muscle - rabbit  
 C:Species: Oryctolagus cuniculus (domestic rabbit)

C>Date: 30-Apr-1979 #sequence\_revision 30-Apr-1979 #text\_change 09-Jul-2004  
 C:Accession: A03089  
 R:Grand, R.J.A.; Wilkinson, J.M.

Biochem. J. 167, 183-192, 1977  
 A:Title: The amino acid sequence of rabbit slow-muscle troponin I.

A:Reference number: A90296; MUID:78060292; PMID:588250  
 A:Accession: A03089  
 A:Molecule type: protein  
 A:Residues: 1-184 <GRA>

A:Cross-references: UNIPROT:P02645  
 A>Note: some of the molecules lack residues 183 and 184  
 C:Complex: troponin is a heterotrimer with one molecule each of troponin C (calcium bind

C:Function:  
 A:Description: binds actin and inhibits myosin ATPase activity; with tropomyosin mediate

C:Superfamily: troponin I  
 A:Pathway: muscle contraction

C:Keywords: actin binding; blocked amino end  
 F:1/Modified site: blocked amino end (Pro) (partial) (probably acetylated) #status exper

Query Match 42.9%; Score 6; DB 1; Length 184;  
 Best Local Similarity 100.0%; Pred. No. 37;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 RRSKIT 8  
 Db 5 RRSKIT 10

## RESULT 7

B44786  
 troponin I, slow skeletal muscle - rat

C:Species: Rattus norvegicus (Norway rat)

C>Date: 19-Mar-1993 #sequence\_revision 19-Mar-1993 #text\_change 09-Jul-2004  
 C:Accession: B44786

R:Koppe, R.I.; Hallauer, P.L.; Karpaci, G.; Hastings, K.E.M.

J. Biol. Chem. 264, 14327-14333, 1989  
 A:Title: cDNA clone and expression analysis of rodent fast and slow skeletal muscle tropo

A:Reference number: A44786; MUID:89340548; PMID:2760067  
 C:Accession: B44786

A:Status: preliminary  
 A:Molecule type: mRNA

A:Residues: 1-187 <KOP>  
 A:Cross-references: UNIPROT:P13413; GB:J04993; NID:g207517; PIDN:AAA42295.1; PID:g207518

C:Superfamily: troponin I  
 C:Keywords: skeletal muscle

Query Match 42.9%; Score 6; DB 2; Length 187;  
 Best Local Similarity 100.0%; Pred. No. 38;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 RRSKIT 8  
 Db 6 RRSKIT 11

## RESULT 8

C49600  
 coat protein 22k - soybean dwarf virus

C:Species: soybean dwarf virus  
 C>Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 09-Jul-2004  
 C:Accession: C49600

R:Rathjen, J.P.; Karsgeorgos, L.E.; Habili, N.; Waterhouse, P.M.; Symons, R.H.  
 Virology 198, 671-679, 1994

A:Title: Soybean dwarf luteovirus contains the third variant genome type in the luteovir  
 A:Reference number: A49600; MUID:94120742; PMID:8291248  
 A:Accession: C49600

A:Status: preliminary  
 A:Molecule type: mRNA

A:Residues: 1-200 <RAV>  
 A:Cross-references: UNIPROT:O87033; GB:I24049; NID:g436017; PIDN:AAA17537.1; PID:g436020

C:Superfamily: potato leaf roll virus coat protein; potato leaf roll virus coat protein  
 F:1-200/Domain: potato leaf roll virus coat protein homology <COP>

Query Match 42.9%; Score 6; DB 2; Length 200;  
 Best Local Similarity 100.0%; Pred. No. 40;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 6 KITSVL 11  
 Db 104 KITSVL 109

## RESULT 9

T25853

hypothetical protein T01B11.7 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans  
 C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
 C:Accession: T25853

R:Geisel, C.; Stellyes, L.  
 submitted to the EMBL Data Library, December 1996

A:Description: The sequence of C. elegans cosmid T01B11.  
 A:Reference number: Z20099  
 A:Accession: T25853

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA  
 A:Residues: 1-255 <GET>  
 A:Cross-references: UNIPROT:Q9N9F1, EMBL:U80931, PIDDN:AB38004.1, GSPDB:GN00022, CESP:TD  
 A:Experimental source: Strain Bristol N2; clone T01B1  
 C:Genetics:  
 A:Gene: CESP:T01B1.7  
 A:Map position: 4  
 A:introns: 72/1; 114/3; 154/3; 195/2; 214/3

Query Match 42.9%; Score 6; DB 2; Length 255;  
 Best Local Similarity 100.0%; Pred. No. 48;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Py 8 TSVLSF 13  
 Db 246 TSVLSF 251

RESULT 10  
 D89830  
 hypothetical protein SA0567 [imported] - Staphylococcus aureus (strain N315)  
 C:Species: Staphylococcus aureus  
 C>Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 09-Jul-2004  
 C:Accession: D89830  
 R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguma, A.; Mizutani, H.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiratake, K.  
 C:Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiratake, K.  
 A>Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.  
 A:Reference number: A89758; MUID:21311952; PMID:11418146  
 A:Accession: D89830  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-316 <KOR>  
 A:Cross-references: UNIPROT:Q99W02, GB:BA000018, PID:g13700502, PIDDN:BB41799.1, GSPDB:G  
 A:Experimental source: strain N315  
 C:Genetics:  
 A:Gene: SA0567  
 C:Superfamily: vitamin B12 transport protein btuc

Query Match 42.9%; Score 6; DB 2; Length 316;  
 Best Local Similarity 100.0%; Pred. No. 58;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Py 7 ITSVLS 12  
 Db 118 ITSVLS 123

RESULT 11  
 S70390  
 DMCL/LIM15 homolog 1 - yeast (Candida albicans)  
 C:Species: Candida albicans  
 C>Date: 06-Dec-1996 #sequence\_revision 13-Mar-1997 #text\_change 09-Jul-2004  
 C:Accession: S70390  
 R:Diener, A.C.; Fink, G.R.  
 C:Diener, A.C.; Fink, G.R.  
 A>Title: LIM15 is a functional Candida albicans homologue of the meiosis-specific gene DMC1.  
 A:Reference number: S70390; MUID:96363911; PMID:8725225  
 A:Accession: S70390  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-324 <DIE>  
 A:Cross-references: UNIPROT:P50265, EMBL:U39808, NID:g1145715, PIDDN:AAC49400.1, PID:g114  
 C:Genetics:  
 A:Gene: DLM1  
 A:introns: 35/3; 67/3  
 C:Superfamily: yeast DNA repair protein RAD51

Query Match 42.9%; Score 6; DB 2; Length 324;  
 Best Local Similarity 100.0%; Pred. No. 59;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Py 7 ITSVLS 12  
 Db 34 ITSVLS 39

RESULT 12  
 T5181  
 protein farnesyltransferase (RC 2.5.1.-) chain A [imported] - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C>Date: 18-Aug-2000 #sequence\_revision 18-Aug-2000 #text\_change 09-Jul-2004  
 C:Accession: T5181  
 R:Belbin, L.; Villarejo, R.; Inze, D.; Thomas, D.; Thomasset, B.  
 A>Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
 A:Reference number: Z55461  
 A:Accession: T5181  
 A:Status: preliminary; translated from GB/EMBL/DDBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-326 <BEL>  
 A:Cross-references: UNIPROT:Q9LX33, EMBL:AF064542, PIDDN:AAC61853.1  
 C:Genetics:  
 A:Gene: FTA  
 C:Keywords: transferase

Query Match 42.9%; Score 6; DB 2; Length 326;  
 Best Local Similarity 100.0%; Pred. No. 59;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Py 3 RKSKIT 8  
 Db .317 RKSKIT 322

RESULT 13  
 T00797  
 hypothetical protein At2g32710 [imported] - Arabidopsis thaliana  
 N:Alternate names: hypothetical protein F24L7.15  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C>Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 09-Jul-2004  
 C:Accession: T00797; E84736  
 R:Rounsley, S.D.; Lin, X.; Kechum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul, M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.; euse, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
 A>Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
 A:Reference number: Z14204  
 A:Accession: T00797  
 A:Status: translated from GB/EMBL/DDBJ  
 A:Molecule type: DNA  
 A:Residues: 1-327 <ROU>  
 A:Cross-references: UNIPROT:O48846, EMBL:AC003974, NID:g2914688, PID:g2914702  
 A:Experimental source: cultivar Columbia  
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Bentto, M.I.; Town, C.D.; Fujii, C.Y.; M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.; euse, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
 A>Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
 A:Reference number: A84420; MUID:20083487; PMID:10617197  
 A:Accession: E84736  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-327 <STO>  
 A:Cross-references: GB:AE002093, NID:g2914702, PIDDN:AAC04492.1, GSPDB:GN00139  
 C:Genetics:  
 A:Gene: At2g32710; F24L7.15  
 A:Map position: 2  
 A:introns: 193/2  
 C:Superfamily: Arabidopsis thaliana hypothetical protein F24L7.15

Query Match 42.9%; Score 6; DB 2; Length 327;  
 Best Local Similarity 100.0%; Pred. No. 60;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Py 2 IRKSKI 7  
 Db 5 IRKSKI 10

## RESULT 14

T49006  
 Larnesyltransferase subunit A (PTA) - Arabidopsis thaliana  
 N/Alternate names: protein F25L23.240  
 C/Species: Arabidopsis thaliana (mouse-ear cress)  
 C/Date: 02-Jun-2000 #sequence\_revision 02-Jun-2000 #text\_change 09-Jul-2004  
 C/Accession: T49006  
 R/D Angelo, M.; Vezzi, A.; Modesto, D.; Pigazzi, M.; Valle, G.; Mewes, H.W.; Rudd, S.; L submitted to the Protein Sequence Database, May 2000  
 A/Reference number: 225012  
 A/Accession: T49006  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-340 <DNA>  
 A/Cross-references: UNIPROT:O9LX33; EMBL:AL356014; GSPDB:GN00061; ATSP:F25L23.240  
 A/Experimental source: cultivar Columbia; BAC clone F25L23  
 C/Genetics:  
 A/Map position: 3  
 A/Genes: ATSP:F25L23.240  
 A/Introns: 89/3; 126/2; 166/3; 198/3

Query Match 42.9%; Score 6; DB 2; Length 340;  
 Best Local Similarity 100.0%; Pred. No. 61;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 RKSXIT 8  
 |||||  
 Db 331 RKSXIT 336

## RESULT 15

A75597  
 mannosyltransferase - Deinococcus radiodurans (strain R1)  
 C/Species: Deinococcus radiodurans  
 C/Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
 C/Accession: A75597  
 R/White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; M.; Shen, M.; Vanathavan, J.U.; Lam, P.; McDonald, L.; Uterback, T.; Zalewski, C.; Ma S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
 Science 286, 1571-1577, 1999  
 A/Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
 A/Reference number: A75250; MUID:20036896; PMID:10567266  
 A/Accession: A75597  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-343 <MHI>  
 A/Cross-references: UNIPROT:O9R2B5; GB:AE001662; GB:AE001825; NID:g6460468; PIDN:AAF1227  
 A/Experimental source: strain R1  
 C/Genetics:  
 A/Map position: 2  
 A/Genes: DRA0039

Query Match 42.9%; Score 6; DB 2; Length 343;  
 Best Local Similarity 100.0%; Pred. No. 62;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ITSVLS 12  
 |||||  
 Db 24 ITSVLS 29

## RESULT 16

D95233  
 aminotransferase, class I [imported] - Streptococcus pneumoniae (strain TIGR4)  
 C/Species: Streptococcus pneumoniae  
 C/Date: 03-Aug-2001 #sequence\_revision 03-Aug-2001 #text\_change 09-Jul-2004  
 C/Accession: D95233  
 R/Tettein, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid on, J.D.; Unayam, L.A.; White, O.; Salberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapfe, nson, T.; Hickey, E.K.; Holt, I.E.  
 Science 293, 498-506, 2001

A/Authors: Lofcus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison, A/Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.  
 A/Reference number: A95000; MUID:21357209; PMID:11463916

A/Accession: D95233  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-404 <KUD>  
 A/Cross-references: UNIPROT:O97NM6; GB:AE005672; PIDN:AAK76061.1; PID:g14973502; GSPDB:G A/Experimental source: strain TIGR4  
 C/Genetics:  
 A/Map position: 3  
 A/Genes: SPI994  
 C/Superfamily: aspartate transaminase

Query Match 42.9%; Score 6; DB 2; Length 404;  
 Best Local Similarity 100.0%; Pred. No. 71;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 KSKITS 9  
 |||||  
 Db 162 KSKITS 167

## RESULT 17

F98097  
 aspartate transaminase (EC 2.6.1.1) [imported] - Streptococcus pneumoniae (strain R6)  
 C/Species: Streptococcus pneumoniae  
 C/Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 09-Jul-2004  
 C/Accession: F98097  
 R/Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; E e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M y, P.; Sun, P.M.; Winkler, M.E.  
 J. Bacteriol. 183, 5709-5717, 2001  
 A/Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaekunas, S.R.; A/Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.  
 A/Reference number: A97872; MUID:2143245; PMID:11544234  
 A/Accession: F98097  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-404 <KUD>  
 A/Cross-references: UNIPROT:O8DNC7; GB:AE007317; PIDN:AAL00611.1; PID:g15459494; GSPDB:G A/Genetics:  
 A/Map position: 2  
 A/Genes: aspc  
 C/Superfamily: aspartate transaminase  
 C/Keywords: aminotransferase

Query Match 42.9%; Score 6; DB 2; Length 404;  
 Best Local Similarity 100.0%; Pred. No. 71;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 KSKITS 9  
 |||||  
 Db 162 KSKITS 167

## RESULT 18

P86240  
 hypothetical protein [imported] - Arabidopsis thaliana  
 C/Species: Arabidopsis thaliana (mouse-ear cress)  
 C/Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 09-Jul-2004  
 C/Accession: P86240  
 R/Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; ansen, N.F.; Hughes, B.; Huizar, L.  
 Nature 408, 816-820, 2000  
 A/Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C. C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
 A/Authors: Salberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
 A/Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
 A/Reference number: A86141; MUID:21016719; PMID:11130712  
 A/Accession: P86240  
 A/Status: preliminary

A:Molecule type: DNA  
A:Residues: 1-429 <STO>  
A:Cross-references: UNIPROT:Q9SAD0; GB:AE005172; NID:g4874276; P1DN:AA031341.1; GSPDB:GN  
C:Genetics:  
A:Map position: 1

Query Match 42.9%; Score 6; DB 2; Length 429;  
Best Local Similarity 100.0%; Pred. No. 75;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 SKITS 9  
DB 416 SKITS 421

RESULT 19  
DESCRS  
homoserine dehydrogenase (EC 1.1.1.3) - *Bacillus subtilis*  
C:Species: *Bacillus subtilis*  
C>Date: 30-Jun-1991 #sequence\_revision 02-Jul-1998 #text\_change 09-Jul-2004  
C:Accession: A31973; C25364; D69642  
R:Parrot, C.; Cohen, G.N.

J. Biol. Chem. 263, 14654-14660, 1988  
A:Title: Cloning and nucleotide sequence of the *Bacillus subtilis* hom gene coding for ho  
ne dehydrogenases I and II.  
A:Reference number: A31973; MUID:89008330; PMID:3139660  
A:Accession: A31973  
A:Molecule type: DNA  
A:Residues: 1-401, 'O', 403-433 <PAR1>  
A:Cross-references: UNIPROT:P1982; EMBL:M23217; NID:g340846; P1DN:AAA50609.1; PID:g5584  
A:Note: These authors used TTG as a start codon; the codon given for 402-Glu (CAA) is in  
R:Parrot, C.

EMBO J. 5, 3013-3019, 1986  
A:Title: Evolution of biosynthetic pathways: a common ancestor for threonine synthase, t  
A:Reference number: A91055; MUID:87080286; PMID:3098560  
A:Accession: C25364  
A:Molecule type: DNA  
A:Residues: 353-374, 'T', 376-433 <PAR2>  
A:Cross-references: GB:X04603; NID:g40210; P1DN:CAA28269.1; PID:8009663  
R:Kunec, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bester  
C.; Bron, S.; Brouillette, S.; Brusch, C.V.; Caldwell, B.; Capiano, V.; Carey, N.M.; Cho  
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrati, E.  
Nature 350, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Funa, S.; Galizzi, A.; Gallier  
leech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.  
Koetter, P.; Koningsstein, G.; Krogsh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinis  
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel  
Y. M.; Ogawa, K.; Ogiwara, A.; Oudaga, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle  
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadai, Y.; Sato, T.; Scanlon,  
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Serot  
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpestre, P.; Tognoni, A.; Tosato, V.; Uchiyama,  
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K  
A:Authors: Yoshikawa, H.F.; Zumeitein, E.; Yoshikawa, H.; Danchin, A.  
A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.  
A:Reference number: A69580; MUID:98044033; PMID:9384377  
A:Accession: D69642  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-433 <KUN>  
A:Cross-references: GB:Z99120; GB:AL009126; NID:g2635613; P1DN:CAB15216.1; PID:g2635723  
A:Experimental source: strain 168  
C:Genetics:  
A:Gene: hom  
A:Map position: 78 min  
C:Function:

A:Description: catalyzes the reduction of L-aspartate 4-semialdehyde to L-homoserine by  
A:Pathway: threonine biosynthesis

C:Superfamily: homoserine dehydrogenase; homoserine dehydrogenase homology  
C:Keywords: NADP, oxidoreductase; threonine biosynthesis  
F:1-243/Domain: homoserine dehydrogenase homology <HSD>  
F:5-33/Region: beta-alpha-beta NADP nucleotide-binding fold

Query Match 42.9%; Score 6; DB 1; Length 433;

Best Local Similarity 100.0%; Pred. No. 75;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 SKITSV 10  
DB 363 SKITSV 368

RESULT 20  
A83728

adenylosuccinate lyase purB [imported] - *Bacillus halodurans* (strain C-125)  
C:Species: *Bacillus halodurans*  
C>Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 09-Jul-2004  
C:Accession: A83728

R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira  
Nucleic Acids Res. 28, 4317-4331, 2000  
A:Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and  
A:Reference number: A83650; MUID:20512582; PMID:11058132  
A:Accession: A83728  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-433 <STO>  
A:Cross-references: UNIPROT:Q9KFE1; GB:AB001509; GB:BA000004; NID:g10173176; P1DN:BA043  
A:Experimental source: strain C-125  
C:Genetics:  
A:Gene: purB  
C:Superfamily: fumarate hydratase

Query Match 42.9%; Score 6; DB 2; Length 433;  
Best Local Similarity 100.0%; Pred. No. 75;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ITSVLS 12  
DB 399 ITSVLS 404

RESULT 21  
D84949

NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain N [imported] - *Buchnera* sp. (strain  
C:Species: *Buchnera* sp.  
C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 03-Jun-2002  
C:Accession: D84949

R:Shigenobu, S.; Watanabe, H.; Hattori, M.; Sakaki, Y.; Ishikawa, H.  
Nature 407, 81-86, 2000  
A:Title: Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. A  
A:Reference number: A84930; MUID:20445173; PMID:10993077  
A:Accession: D84949  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-469 <STO>  
A:Cross-references: GB:AP000398; GSPDB:GN00144  
A:Experimental source: strain APS  
C:Genetics:

A:Gene: nuon; BUI66  
C:Superfamily: NADH dehydrogenase (ubiquinone) chain 2  
C:Keywords: NAD, oxidoreductase  
Query Match 42.9%; Score 6; DB 2; Length 469;  
Best Local Similarity 100.0%; Pred. No. 80;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 TSVLSF 13  
DB 220 TSVLSF 225

RESULT 22  
G86227

hypothetical protein [imported] - *Arabidopsis thaliana*  
C:Species: *Arabidopsis thaliana* (mouse-ear cress)  
C>Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 09-Jul-2004  
C:Accession: G86227

R:Theologos, A.; Ecker, J.R.; Palm, C.J.; Federpiel, N.A.; Kaul, S.; White, O.; Alonso, C.M.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; Jensen, N.F.; Hughes, B.; Hultzer, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lucero, J.S.; Maiti, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, Ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant *Arabidopsis*.  
A:Reference number: A6141; MUID:21016719; PMID:11130712  
A:Accession: G86227  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-482 <STO>  
A:Cross-references: UNIPROT:O80527; GB:AE005172; NID:g3482919; PIDN:AAC33204.1; GSPDB:GN C:Genetics:  
A:Map position: 1

Query Match 42.9%; Score 6; DB 2; Length 482;  
Best Local Similarity 100.0%; Pred. No. 82;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 SVLSFC 14  
|||||  
Db 23 SVLSFC 28

RESULT 23  
S45492  
1aps protein - fission yeast (*Schizosaccharomyces pombe*)  
C:Species: *Schizosaccharomyces pombe*  
C:Date: 10-Dec-1994 #sequence\_revision 21-Jul-1995 #text\_change 09-Jul-2004  
A:Accession: S45492; S35896  
R:Sato, S.; Suzuki, H.; Widyastuti, U.; Hotta, Y.; Tabata, S.  
Curr. Genet. 26, 31-37, 1994  
A:Title: Identification and characterization of genes induced during sexual differentiation  
A:Reference number: S45492; MUID:95042833; PMID:7954893  
A:Accession: S45492  
A:Status: nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 1-543 <SAT>  
A:Cross-references: UNIPROT:P40901; EMBL:D14062; NID:g218544; PIDN:BA03148.1; PID:g2185  
C:Genetics:  
A:Gene: 1aps  
C:Superfamily: arginine permease  
C:Keywords: transmembrane protein

Query Match 42.9%; Score 6; DB 2; Length 543;  
Best Local Similarity 100.0%; Pred. No. 91;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 SVLSFC 14  
|||||  
Db 537 SVLSFC 542

RESULT 24  
C84632  
hypothetical protein At2g24070 [imported] - *Arabidopsis thaliana*  
C:Species: *Arabidopsis thaliana* (mouse-ear cress)  
C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 09-Jul-2004  
A:Accession: C84632  
R:Lin, X.; Kaul, S.; Rounsaey, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon, L.; Jensen, D.; Mierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
Nature 402, 761-768, 1999  
A:Title: Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*.  
A:Reference number: A84420; MUID:20083487; PMID:10617197  
A:Accession: C84632  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-560 <STO>

A:Cross-references: UNIPROT:O82235; GB:AE002093; NID:g3738334; PIDN:AAC63675.1; GSPDB:GN C:Genetics:  
A:Gene: At2g24070  
A:Map position: 2

Query Match 42.9%; Score 6; DB 2; Length 560;  
Best Local Similarity 100.0%; Pred. No. 93;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 TSULSF 13  
|||||  
Db 365 TSULSF 370

RESULT 25  
S19063  
hypothetical protein YNR023w precursor, mitochondrial - yeast (*Saccharomyces cerevisiae*)  
N:Alternate names: hypothetical protein N3224  
C:Species: *Saccharomyces cerevisiae*  
C:Date: 16-Sep-1992 #sequence\_revision 16-Sep-1992 #text\_change 09-Jul-2004  
A:Accession: S19063; S63354  
R:Donald, K.A.G.; Hill, J.; Griffiths, D.E.  
submitted to the EMBL Data Library, October 1991  
A:Reference number: S19063  
A:Accession: S19063  
A:Molecule type: DNA  
A:Residues: 1-566 <DON>  
A:Cross-references: UNIPROT:P53628; EMBL:X62430; NID:g4106; PID:g4107  
R:Pohl, T.M.  
submitted to the Protein Sequence Database, April 1996  
A:Reference number: S63346  
A:Accession: S63354  
A:Molecule type: DNA  
A:Residues: 1-566 <POH>  
A:Cross-references: EMBL:Z71638; NID:g1302511; PID:e2339821; PID:g1302512; MIPS:YNR023w  
A:Experimental source: strain S288C  
C:Genetics:  
A:Gene: SGD:SNF12  
A:Cross-references: SGD:S0005306; MIPS:YNR023w  
A:Map position: 14R  
A:Genome: nuclear  
C:Keywords: mitochondrion

Query Match 42.9%; Score 6; DB 2; Length 566;  
Best Local Similarity 100.0%; Pred. No. 94;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ITSULS 12  
|||||  
Db 487 ITSULS 492

RESULT 26  
AF0211  
probable ABC transporter (ATP-binding protein) YPO1735 [imported] - *Yersinia pestis* (str  
C:Species: *Yersinia pestis*  
C:Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 09-Jul-2004  
A:Accession: AF0211  
R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Tittball, R.W.; Holden, M.T.G.; Prentice, M.B.  
deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;  
H., M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall,  
Nature 413, 523-527, 2001  
A:Title: Genome sequence of *Yersinia pestis*, the causative agent of plague.  
A:Reference number: AB0001; MUID:21470413; PMID:11586360  
A:Accession: AF0211  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-572 <KUP>  
A:Cross-references: UNIPROT:O8ZFH4; GB:AL590842; PIDN:CAC90554.1; PID:g15979763; GSPDB:G  
A:Gene: YPO1735  
C:Superfamily: Escherichia coli ABC transporter mdla; ATP-binding cassette homology

Query Match 42.9%; Score 6; DB 2; Length 572;  
Best Local Similarity 100.0%; Pred. No. 94;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Cy 6 KITSVL 11  
Db 289 KITSVL 294

## RESULT 27

TS0059  
sexual differentiation process probable amino-acid permease isps [imported] - fission ye  
C:Species: Schizosaccharomyces pombe  
C:Date: 09-Jun-2000 #sequence\_revision 09-Jun-2000 #text\_change 09-Jul-2004  
C:Accession: TS0059  
R: Hunt, C.; Aves, S.; McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.  
submitted to the EMBL Data Library, December 1999  
A:Reference number: Z25031  
A:Accession: TS0059  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Accession: TS0059  
A:Molecule type: DNA  
A:Residues: 1-580 <HUN>  
A:Cross-references: UNIPROT:P40901; EMBL:AL133521; PIDN:CAB63545.1; GSPDB:GN00066; SPDB:  
A:Experimental source: strain 972h(-); cosmid C1039  
C:Genetics:  
A:Gene: SPDB:SPAC1039.09  
A:Map position: 1  
C:Superfamily: arginine permease

Query Match 42.9%; Score 6; DB 2; Length 580;  
Best Local Similarity 100.0%; Pred. No. 96;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 9 SVLSFC 14  
Db 574 SVLSFC 579

## RESULT 28

SS8319  
hypoherical protein YOR308C - Yeast (Saccharomyces cerevisiae)  
N:Alternate names: hypothetical protein O567  
C:Species: Saccharomyces cerevisiae  
C:Date: 13-Jan-1996 #sequence\_revision 01-Mar-1996 #text\_change 09-Jul-2004  
C:Accession: SS8319; S67214; S71987  
R: Pearson, B.M.; Hernando, Y.; Wolf, S.S.; Kalogeropoulos, A.; Schweizer, M.  
submitted to the EMBL Data Library, August 1995  
A:Reference number: S58318  
A:Accession: SS8319

A:Molecule type: DNA  
A:Residues: 1-587 <PEA>  
A:Cross-references: UNIPROT:Q12420; EMBL:X90565; NID:g940836; PID:g940838  
R: Pearson, B.M.; Hernando, Y.; Kalogeropoulos, A.; Schweizer, M.  
submitted to the Protein Sequence Database, July 1996  
A:Reference number: S67213  
A:Accession: S67214

A:Molecule type: DNA  
A:Residues: 1-587 <PEA>  
A:Cross-references: EMBL:Z75216; NID:g1420678; PID:g252142; PID:g1420679; MIPS:YOR308C  
R: Czeplich, C.; Jantiaux, J.C.; Korde, E.; Polrey, R.; Pujol, A.; Tobiasch, E.  
submitted to the Protein Sequence Database, July 1996  
A:Reference number: S67194  
A:Accession: S67212

A:Accession: S67212

A:Residues: 285-587 <CZI>  
A:Cross-references: EMBL:Z75216; MIPS:YOR308C

A:Experimental source: strain S288C

R: Pearson, B.M.; Hernando, Y.; Payne, J.; Wolf, S.S.; Kalogeropoulos, A.; Schweizer, M.  
Yeast 12, 1021-1031, 1996

A:Title: Sequencing of a 35.71 kb DNA segment on the right arm of yeast chromosome XV re  
A:Reference number: S71986; MUID:97051589; PMID:8896266

A:Accession: S71987

A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-587 <PEA>  
A:Cross-references: EMBL:X90565; NID:g940836; PIDN:CAB62162.1; PID:g940838  
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1995  
C:Genetics:  
A:Gene: SGD:SNUG6  
A:Cross-references: SGD:S0005835  
A:Map position: 15R

Query Match 42.9%; Score 6; DB 2; Length 587;  
Best Local Similarity 100.0%; Pred. No. 97;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 3 RKSKIT 8  
Db 294 RKSKIT 299

## RESULT 29

D85359  
hypoherical protein AT4g30710 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 09-Jul-2004  
C:Accession: D85359  
R: anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring  
Nature 402, 769-777, 1999  
A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.  
A:Reference number: A85001; MUID:20083488; PMID:10617198  
A:Accession: D85359

A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-644 <STO>  
A:Cross-references: UNIPROT:Q9SUH5; GB:NC\_001268; NID:g726972; PIDN:CAB79789.1; GSPDB:G  
A:Gene: AT4g30710  
A:Map position: 4

Query Match 42.9%; Score 6; DB 2; Length 644;  
Best Local Similarity 100.0%; Pred. No. 1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 8 TSVLSF 13  
Db 410 TSVLSF 415

## RESULT 30

G64068  
DNA topoisomerase (EC 5.99.1.2) III - Haemophilus influenzae (strain Rd KW20)  
C:Species: Haemophilus influenzae  
C:Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 09-Jul-2004  
C:Accession: G64068  
R: Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A  
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodde, A.; Kelley, J.M.; Weidman, J  
; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghegan, N.S.M.  
Science 269, 496-512, 1995  
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, A  
A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.  
A:Reference number: A64000; MUID:95350630; PMID:7542800

A:Accession: G64068  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-651 <TRGR>  
A:Cross-references: UNIPROT:P43704; GB:U32727; GB:U42023; NID:g1573415; PIDN:AAC22103.1;  
C:Superfamily: DNA topoisomerase I topa

C:Keywords: DNA binding; isomerase

Query Match 42.9%; Score 6; DB 2; Length 651;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 3 RKSKIT 8



Db 431 RKSKIT 436

## RESULT 31

H86221

hypothetical protein [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 09-Jul-2004

C:Accession: H86221

R:Theologian: A.; Ecker, J.R.; Palm, C.J.; Federpiel, N.A.; Kaul, S.; White, O.; Alonso, C.A.; Li, J.H.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Matli, R.; Marziani, R.; Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salberg, S.L.; Schwartz, J.R.; Shim, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719; PMID:11130712

A:Accession: H86221

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-652 &lt;STO&gt;

A:Cross-references: UNIPROT:O04030; GB:AE005172; NID:G2342692; PIDN:AB70419.1; GSPDB:GN

C:Genetics:

A:Map position: 1

Query Match 42.9%; Score 6; DB 2; Length 652;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 IRKSKI 7  
|||||  
Db 41 IRKSKI 46

## RESULT 32

AC0323

ribonucleoside-diphosphate reductase (EC 1.17.4.1) alpha chain [imported] - Yersinia pes

C:Species: Yersinia pestis

C:Date: 02-Nov-2001 #sequence\_revision 02-Nov-2001 #text\_change 09-Jul-2004

C:Accession: AC0323

R:Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titchell, R.W.; Holden, M.T.G.; Prentice, M.B.

deno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;

11, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrall, Nature 413, 523-527, 2001

A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.

A:Reference number: AB0001; MUID:21470413; PMID:11586360

A:Accession: AC0323

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-693 &lt;KUN&gt;

A:Cross-references: UNIPROT:Q8ZDC7; GB:AL590842; PIDN:CA02890.1; PID:G15980632; GSPDB:G

C:Genetics:

A:Gene: nrdB

C:Superfamily: Salmonella typhimurium ribonucleoside-diphosphate reductase

C:Keywords: oxidoreductase

Query Match 42.9%; Score 6; DB 2; Length 693;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 IRKSKI 7  
|||||  
Db 319 IRKSKI 324

## RESULT 33

T23585

hypothetical protein K1004.4 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004

C:Accession: T23585

R:Steward, C.

submitted to the EMBL Data Library, March 1997

A:Reference number: 219764

A:Accession: T23585

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-727 &lt;WIL&gt;

A:Cross-references: UNIPROT:O45675; EMBL:Z92806; PIDN:CA07259.1; GSPDB:GN00023; CESP:KI

A:Experimental source: clone K1004

C:Genetics:

A:Gene: CESP:K1004.4

A:Map position: 5

A:Introns: 168/3; 259/2; 494/3; 555/3

Query Match 42.9%; Score 6; DB 2; Length 727;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 TSVLSP 13  
|||||  
Db 415 TSVLSP 420

## RESULT 34

C96552

hypothetical protein F5D21.12 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 09-Jul-2004

C:Accession: C96552

R:Theologian: A.; Ecker, J.R.; Palm, C.J.; Federpiel, N.A.; Kaul, S.; White, O.; Alonso, C.A.; Li, J.H.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Matli, R.; Marziani, R.; Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salberg, S.L.; Schwartz, J.R.; Shim, P.; Southwick, A.M.; Sun, H.; Tallon, ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719; PMID:11130712

A:Accession: C96552

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-809 &lt;STO&gt;

A:Cross-references: UNIPROT:Q9C8J3; GB:AE005173; NID:G10092348; PIDN:AG12757.1; GSPDB:G

C:Genetics:

A:Gene: F5D21.12

A:Map position: 1

Query Match 42.9%; Score 6; DB 2; Length 809;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MIRSK 6  
|||||  
Db 9 MIRSK 14

## RESULT 35

T41645

probable spindle pole body-associated protein - fission yeast (Schizosaccharomyces pombe

C:Species: Schizosaccharomyces pombe

C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004

C:Accession: T41645

R:Wood, V.; Rajandream, M.A.; Barrall, B.G.; Lucas, M.; Galliard, C.

submitted to the EMBL Data Library, January 1999

A:Reference number: Z22006

A:Accession: T41645

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-809 &lt;WOO&gt;

A:Cross-references: UNIPROT:Q94534; EMBL:AL035247; PDB:CAA22843.1; GSPDB:GN00068; SPDB:  
A:Experimental source: strain 972h; cosmid c895  
C:Genetics:  
A:Gene: SPDB:SPCC895.07  
A:Map position: 3

Query Match 42.9%; Score 6; DB 2; Length 809;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 IRKSKI 7  
Db 491 IRKSKI 496

RESULT 36  
T33217  
hypochemical protein T07H8.6 - *Caenorhabditis elegans*

C:Species: *Caenorhabditis elegans*  
C>Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T33217  
R:Gating, S.; Maggi, L.  
submitted to the EMBL Data Library, May 1998  
A:Description: The sequence of *C. elegans* cosmid T07H8.  
A:Reference number: 221303  
A:Accession: T33217  
A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA  
A:Residues: 1-840 <GAT>  
A:Cross-references: UNIPROT:O61894; EMBL:AF067945; PDB:AACT1680.1; GSPDB:GN00023; CESP:  
A:Experimental source: strain Bristol N2; clone T07H8  
C:Genetics:  
A:Gene: CESP:T07H8.6  
A:Map position: 5  
A:Introns: 55/3; 86/2; 130/3; 162/3; 217/3; 281/2; 475/3; 543/3; 586/2; 617/3; 692/1; 73

Query Match 42.9%; Score 6; DB 2; Length 840;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 TSVLSF 13  
Db 375 TSVLSF 380

RESULT 37

149583  
differentiation antigen - mouse  
C:Species: *Mus musculus* (house mouse)  
C>Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 09-Jul-2004  
C:Accession: 149583  
R:Law, C.

U: Immunol. 151, 175-187, 1993  
A:Title: Organization of the murine Cd32 locus. Mapping to chromosome 7 and characteriza  
A:Reference number: 149583; MUID:93115834; PMID:8100843  
A:Accession: 149583

A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-862 <RES>  
A:Cross-references: UNIPROT:P35329; GB:LI6928; NID:G348965; PDB:AAA02562.1; PID:G348966  
C:Genetics:  
A:Gene: CD22

Query Match 42.9%; Score 6; DB 2; Length 862;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 SKITSV 10  
Db 185 SKITSV 190

RESULT 38

669599

acetylase hydratase (EC 4.2.1.3) citB - *Bacillus subtilis*

N:Alternate names: acetylase

C:Species: *Bacillus subtilis*

C>Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 09-Jul-2004

C:Accession: 669599; A27085

R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertier  
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Cho  
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.  
Nature 390, 249-256, 1997

A:Authors: Poulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizi, A.; Gallier  
lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holtsappel, S.; Hosono, S.; Hullo, M.F.  
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,  
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel,  
Y. M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portecelle  
Rieger, M.; Rivolta, C.; Roche, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon,  
A:Authors: Schleicher, S.; Schrotter, R.; Scorfione, F.; Sekiguchi, J.; Sekowska, A.; Seror  
akenuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpestra, P.; Tognoni, A.; Tosato, V.; Uchiyama,  
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K  
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.  
A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.  
A:Reference number: A69580; MUID:98044033; PMID:9384377

A:Accession: 669599

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-909 <KUN>

A:Cross-references: UNIPROT:P09339; GB:Z29113; GB:AL009126; NID:G2634090; PDB:CAB13684.1

A:Experimental source: strain 168

R: Dingman, D.W.; Sonenshein, A.L.

U: Bacteriol. 169, 3062-3067, 1987

A:Title: Purification of acetylase from *Bacillus subtilis* and correlation of its N-termir  
A:Reference number: A27085; MUID:87250270; PMID:3110133

A:Accession: A27085

A:Molecule type: DNA

A:Residues: 1-28, 'I', 30-42, 877-879 <DIN>

A>Note: the authors translated the codon ATC for residue 29 as Tyr

C:Genetics:

A:Gene: citB

C:Function:

A:Description: reversibly catalyzes the hydration of cis-acetate to citrate and also tr

A:Pathway: tricarboxylic acid cycle

C:Superfamily: iron-responsive element-binding protein

C:Keywords: 4Fe-4S; carbon-oxygen lyase; hydro-lyase; iron-sulfur protein; metalloprotein

F:493-636/Domain: RNA binding #status predicted <RNA>

F:135,136,189,219,221,316,792/Active site: Asp, His, His, Asp, His, Glu, Ser #status pre

F:450,516,519/Binding site: 4Fe-4S cluster (Cys) (covalent) #status predicted

Query Match 42.9%; Score 6; DB 2; Length 909;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ITSVLS 12  
Db 541 ITSVLS 546

RESULT 39  
T27053

hypochemical protein Y49E10.19 - *Caenorhabditis elegans*

C:Species: *Caenorhabditis elegans*

C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004

C:Accession: T27053

R:Barlow, K.

submitted to the EMBL Data Library, August 1997

A:Reference number: Z20303

A:Accession: T27053

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1205 <NTL>

A:Cross-references: UNIPROT:Q9XTT4; EMBL:Z98866; PDB:CAB11565.1; GSPDB:GN00021; CESP:Y4

A:Experimental source: clone Y49E10  
A:Gene: CESP:Y49E10.19

A:Map position: 3  
A:Introns: 43/1; 281/3; 370/3; 418/3; 836/2; 949/3; 995/3; 1119/3

Query Match 42.9%; Score 6; DB 2; Length 1205;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 SKITSV 10  
|||  
Db 34 SKITSV 39

## RESULT 40

S78061

DNA-directed RNA polymerase (EC 2.7.7.6) II - African swine fever virus

N/Alternate names: EPI242L protein

C/Species: African swine fever virus, ASFV

C/Date: 19-Nov-1997 #sequence\_revision 05-Dec-1997 #text\_change 09-Jul-2004

C/Accession: S78061; S33996

R/Yanez, R.J.

submitted to the EMBL Data Library, January 1993

A/Reference number: S78060

A/Accession: S78061

A/Molecule type: DNA

A/Residues: 1-1242 <YAN>

A/Cross-references: UNIPROT:P42487; EMBL:Z21490; NID:g311527; PIDN:CAA79698.1; PID:g3115

R/Yanez, R.J.; Bourneil, M.; Nogai, M.L.; Yuste, L.; Vinuela, E.

Nucleic Acids Res. 21, 2423-2427, 1993

A/Title: African swine fever virus encodes two genes which share significant homology w

A/Reference number: S33995; MUID:93281390; PMID:8506138

A/Accession: S33996

A>Status: nucleic acid sequence not shown

A/Molecule type: DNA

A/Residues: 99-212;367-428;538-679;770-880;910-1048;1074-1201 <YAN>

A/Cross-references: EMBL:Z21490

A/Experimental source: strain BA71V

C/Superfamily: DNA-directed RNA polymerase 132K polypeptide

C/Keywords: DNA binding; nucleosidyltransferase; transcription; zinc finger

Query Match 42.9%; Score 6; DB 2; Length 1242;

Best Local Similarity 100.0%; Pred. No. 1.8e+02;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 ITSVLS 12  
|||  
Db 1230 ITSVLS 1235

Search completed: December 30, 2004, 15:33:59  
Job time : 45 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 30, 2004, 15:27:39 : Search time 37 Seconds  
(without alignments)  
25.093 Million cell updates/sec

Title: US-10-719-385-20  
Perfect score: 14  
Sequence: 1 MIRSKITSVSFC 14

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 478139 seqs, 66318000 residues

Word size : 0

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Listing first 150 summaries

Database : Issued Patents AA:\*

1: /cgn2\_6/ptodata/1/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/1/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/PCITUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	7	50.0	162	4	US-09-134-000C-3462
2	6	42.9	77	4	US-09-513-999C-4730
3	6	42.9	133	4	US-09-107-532A-4535
4	6	42.9	135	4	US-09-543-681A-4409
5	6	42.9	143	4	US-09-732-210-617
6	6	42.9	183	4	US-09-543-681A-7075
7	6	42.9	238	4	US-09-583-110-3590
8	6	42.9	270	4	US-09-107-532A-7334
9	6	42.9	275	4	US-09-248-796A-16584
10	6	42.9	311	4	US-09-489-039A-9786
11	6	42.9	327	4	US-09-526-597D-34
12	6	42.9	344	4	US-09-107-532A-5546
13	6	42.9	433	1	US-08-700-335-20
14	6	42.9	494	4	US-09-019-095A-24
15	6	42.9	540	4	US-08-945-771-2
16	6	42.9	661	4	US-09-371-338-7
17	6	42.9	739	3	US-09-134-001C-3586
18	6	42.9	967	3	US-09-543-681A-6407
19	5	35.7	12	3	US-09-177-249-124
20	5	35.7	13	3	US-09-522-666-30
21	5	35.7	15	1	US-08-268-251-5
22	5	35.7	15	5	PCR-US93-01112-5
23	5	35.7	16	3	US-09-360-237-55
24	5	35.7	16	4	US-09-556-877-251
25	5	35.7	16	4	US-09-620-412C-251
26	5	35.7	16	4	US-09-598-419-251
27	5	35.7	16	4	US-09-680-571A-83

28	5	35.7	16	4	US-09-680-571A-97	Sequence 97, Appl
29	5	35.7	20	4	US-09-556-877-248	Sequence 248, App
30	5	35.7	20	4	US-09-556-877-249	Sequence 249, App
31	5	35.7	20	4	US-09-556-877-250	Sequence 250, App
32	5	35.7	20	4	US-09-620-412C-248	Sequence 248, App
33	5	35.7	20	4	US-09-620-412C-249	Sequence 249, App
34	5	35.7	20	4	US-09-620-412C-250	Sequence 250, App
35	5	35.7	20	4	US-09-598-419-248	Sequence 248, App
36	5	35.7	20	4	US-09-598-419-249	Sequence 249, App
37	5	35.7	20	4	US-09-598-419-250	Sequence 250, App
38	5	35.7	20	4	US-09-834-759-545	Sequence 250, App
39	5	35.7	23	1	US-08-268-251-35	Sequence 35, Appl
40	5	35.7	23	1	PCR-US93-01112-35	Sequence 35, Appl
41	5	35.7	27	3	US-09-360-237-57	Sequence 57, Appl
42	5	35.7	32	4	US-09-759-112A-17	Sequence 17, Appl
43	5	35.7	37	4	US-09-207-388-17	Sequence 17, Appl
44	5	35.7	39	3	US-08-984-277-10	Sequence 10, Appl
45	5	35.7	40	3	US-09-286-691-17	Sequence 17, Appl
46	5	35.7	40	3	US-09-687-147-17	Sequence 17, Appl
47	5	35.7	44	4	US-09-443-780C-19	Sequence 19, Appl
48	5	35.7	44	4	US-09-079-723-3	Sequence 3, Appl
49	5	35.7	50	4	US-09-583-110-4798	Sequence 4798, App
50	5	35.7	51	4	US-09-621-976-7274	Sequence 7274, App
51	5	35.7	53	4	US-09-513-999C-7248	Sequence 7248, App
52	5	35.7	60	4	US-09-513-999C-4510	Sequence 4510, App
53	5	35.7	61	3	US-09-134-001C-4270	Sequence 4270, App
54	5	35.7	63	4	US-09-489-039A-7831	Sequence 7831, App
55	5	35.7	63	4	US-09-489-039A-10388	Sequence 10388, App
56	5	35.7	63	4	US-09-248-796A-25683	Sequence 25683, A
57	5	35.7	63	4	US-09-248-796A-27206	Sequence 27206, A
58	5	35.7	63	4	US-09-513-999C-4708	Sequence 4708, App
59	5	35.7	64	4	US-09-149-476-449	Sequence 449, App
60	5	35.7	64	4	US-09-270-767-54318	Sequence 54318, A
61	5	35.7	64	4	US-09-270-767-54635	Sequence 54635, A
62	5	35.7	65	4	US-09-328-352-5178	Sequence 5178, App
63	5	35.7	67	4	US-09-583-110-5275	Sequence 5275, App
64	5	35.7	68	4	US-09-583-110-4780	Sequence 4780, App
65	5	35.7	68	4	US-09-270-767-39522	Sequence 39522, A
66	5	35.7	68	4	US-09-270-767-54739	Sequence 54739, A
67	5	35.7	68	4	US-09-270-767-54739	Sequence 54739, A
68	5	35.7	69	4	US-09-248-796A-4129	Sequence 4129, A
69	5	35.7	70	3	US-09-367-953B-47	Sequence 47, Appl
70	5	35.7	70	4	US-09-489-039A-12544	Sequence 12544, A
71	5	35.7	75	4	US-09-583-110-2743	Sequence 2743, App
72	5	35.7	77	3	US-09-134-001C-4812	Sequence 4812, App
73	5	35.7	78	4	US-09-248-796A-25922	Sequence 25922, A
74	5	35.7	79	4	US-09-513-999C-5308	Sequence 5308, App
75	5	35.7	81	4	US-09-134-000C-3488	Sequence 3488, App
76	5	35.7	81	4	US-09-621-976-4282	Sequence 4282, App
77	5	35.7	81	4	US-09-270-767-61813	Sequence 61813, A
78	5	35.7	82	4	US-09-270-767-61865	Sequence 61865, A
79	5	35.7	82	4	US-09-270-767-65881	Sequence 65881, A
80	5	35.7	87	4	US-09-178-093B-46	Sequence 46, Appl
81	5	35.7	87	4	US-09-248-796A-73266	Sequence 73266, A
82	5	35.7	91	4	US-09-621-976-5762	Sequence 5762, App
83	5	35.7	92	4	US-09-583-110-3504	Sequence 3504, App
84	5	35.7	100	4	US-09-732-210-451	Sequence 451, App
85	5	35.7	103	4	US-09-191-468-23	Sequence 23, Appl
86	5	35.7	103	4	US-09-191-468-25	Sequence 25, Appl
87	5	35.7	103	4	US-09-191-468-27	Sequence 27, Appl
88	5	35.7	103	4	US-09-191-468-29	Sequence 29, Appl
89	5	35.7	103	4	US-09-191-468-31	Sequence 31, Appl
90	5	35.7	103	4	US-09-191-468-33	Sequence 33, Appl
91	5	35.7	103	4	US-09-191-468-35	Sequence 35, Appl
92	5	35.7	103	4	US-09-191-468-37	Sequence 37, Appl
93	5	35.7	106	4	US-09-732-210-1463	Sequence 1463, App
94	5	35.7	108	1	US-08-436-463-17	Sequence 17, Appl
95	5	35.7	110	1	US-08-466-866-29	Sequence 29, Appl
96	5	35.7	110	3	US-08-685-871-62	Sequence 62, Appl
97	5	35.7	110	3	US-08-469-617-29	Sequence 29, Appl
98	5	35.7	110	3	US-08-469-617-29	Sequence 29, Appl
99	5	35.7	111	4	US-09-270-767-59945	Sequence 59945, A
100	5	35.7	116	4	US-09-270-767-33729	Sequence 33729, A

```

101 5 35.7 116 4 US-09-270-767-48946 Sequence 48946, A
102 5 35.7 121 4 US-09-759-112A-7 Sequence 7, Appl1
103 5 35.7 122 1 US-08-436-463-14 Sequence 14, Appl
104 5 35.7 123 4 US-09-248-796A-17814 Sequence 17814, A
105 5 35.7 124 4 US-09-543-681A-6095 Sequence 6095, Ap
106 5 35.7 125 4 US-09-690-454-155 Sequence 155, App
107 5 35.7 126 4 US-09-732-210-269 Sequence 269, App
108 5 35.7 127 4 US-09-732-210-729 Sequence 729, App
109 5 35.7 127 4 US-09-732-210-737 Sequence 737, App
110 5 35.7 128 4 US-09-583-110-4466 Sequence 4466, Ap
111 5 35.7 128 4 US-09-513-999C-7889 Sequence 7889, Ap
112 5 35.7 129 4 US-09-270-767-33971 Sequence 33971, A
113 5 35.7 129 4 US-09-270-767-49188 Sequence 49188, A
114 5 35.7 131 4 US-09-430-866A-4 Sequence 4, Appl1
115 5 35.7 131 4 US-09-248-796A-23124 Sequence 23124, A
116 5 35.7 132 4 US-09-732-210-589 Sequence 589, App
117 5 35.7 133 4 US-09-252-991A-24089 Sequence 24089, A
118 5 35.7 133 4 US-09-252-991A-29861 Sequence 29861, A
119 5 35.7 140 2 US-08-483-636-4 Sequence 4, Appl1
120 5 35.7 140 2 US-08-483-632-4 Sequence 4, Appl1
121 5 35.7 141 2 US-08-483-636-10 Sequence 10, Appl
122 5 35.7 141 2 US-08-483-632-10 Sequence 10, Appl
123 5 35.7 141 2 US-09-248-796A-19288 Sequence 19288, A
124 5 35.7 142 2 US-08-164-292B-20 Sequence 20, Appl
125 5 35.7 142 3 US-08-845-623-20 Sequence 20, Appl
126 5 35.7 142 3 US-08-815-927-20 Sequence 20, Appl
127 5 35.7 142 3 US-09-103-330-20 Sequence 20, Appl
128 5 35.7 142 3 US-09-435-242-20 Sequence 20, Appl
129 5 35.7 142 4 US-09-270-767-39503 Sequence 39503, A
130 5 35.7 142 4 US-09-270-767-40599 Sequence 40599, A
131 5 35.7 142 4 US-09-270-767-54720 Sequence 54720, A
132 5 35.7 142 4 US-09-270-767-55815 Sequence 55815, A
133 5 35.7 143 1 US-07-710-361-9 Sequence 9, Appl1
134 5 35.7 144 4 US-09-248-796A-14975 Sequence 14975, A
135 5 35.7 145 1 US-08-565-386-15 Sequence 15, Appl
136 5 35.7 149 4 US-09-270-767-39579 Sequence 39579, A
137 5 35.7 149 4 US-09-270-767-54796 Sequence 54796, A
138 5 35.7 150 4 US-09-252-991A-26074 Sequence 26074, A
139 5 35.7 150 4 US-09-621-976-4167 Sequence 4167, Ap
140 5 35.7 151 4 US-09-328-352-6429 Sequence 6429, Ap
141 5 35.7 154 3 US-09-387-418A-11 Sequence 11, Appl
142 5 35.7 155 4 US-09-270-767-33093 Sequence 33093, A
143 5 35.7 155 4 US-09-270-767-48310 Sequence 48310, A
144 5 35.7 156 4 US-09-621-976-4168 Sequence 4168, Ap
145 5 35.7 157 4 US-09-270-767-61769 Sequence 61769, A
146 5 35.7 165 4 US-09-489-039A-11468 Sequence 11468, A
147 5 35.7 167 4 US-09-270-767-44927 Sequence 44927, A
148 5 35.7 174 1 US-08-261-825-2 Sequence 2, Appl1
149 5 35.7 174 2 US-08-719-124-2 Sequence 2, Appl1
150 5 35.7 174 5 PCR-US95-07748A-2 Sequence 2, Appl1

```

## ALIGNMENTS

```

RESULT 1
US-09-134-000C-3462
; Sequence 3462, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3462
; LENGTH: 162
; TYPE: PRT

```

```

; ORGANISM: Enterococcus faecalis
US-09-134-000C-3462

```

```

Query Match 50.0%; Score 7; DB 4; Length 162;
Best Local Similarity 100.0%; Pred. No. 5.1;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 6 KITSVLS 12
Db 129 KITSVLS 135

```

```

RESULT 2
US-09-513-999C-4730
; Sequence 4730, Application US/09513999C
; Patent No. 6783961

```

```

; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J. B.
; APPLICANT: Duclert, A.

```

```

; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6783961

```

```

; FILE REFERENCE: 59.US2.REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24

```

```

; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26

```

```

; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 4730

```

```

; LENGTH: 77
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-513-999C-4730

```

```

Query Match 42.9%; Score 6; DB 4; Length 77;
Best Local Similarity 100.0%; Pred. No. 27;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 4 KSKITS 9
Db 67 KSKITS 72

```

```

RESULT 3
US-09-107-532A-4535
; Sequence 4535, Application US/09107532A
; Patent No. 6583275

```

```

; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS

```

```

; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESS: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA

```

```

; ZIP: 02154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>

```

```

; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998

```

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997

```

```
ATTORNEY/AGENT INFORMATION:
NAME: Ariudello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 4535:
SEQUENCE CHARACTERISTICS:
LENGTH: 133 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...133
SEQUENCE DESCRIPTION: SEQ ID NO: 4535:
US-09-107-532A-4535

Query Match      42.9%; Score 6; DB 4; Length 133;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 KSKITS 9
DB 120 KSKITS 125

RESULT 4
US-09-543-681A-4409
Sequence 4409, Application US/09543681A
Patent No. 6605709
GENERAL INFORMATION:
APPLICANT: GARY BRETON
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
FILE REFERENCE: 2709.1002-001
CURRENT APPLICATION NUMBER: US/09/543, 681A
CURRENT FILING DATE: 2000-04-05
PRIOR APPLICATION NUMBER: US 60/128,706
PRIOR FILING DATE: 1999-04-09
NUMBER OF SEQ ID NOS: 8344
SEQ ID NO 4409
LENGTH: 135
TYPE: PRT
ORGANISM: Proteus mirabilis
US-09-543-681A-4409

Query Match      42.9%; Score 6; DB 4; Length 135;
Best Local Similarity 100.0%; Pred. No. 43;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 KITSVL 11
DB 3 KITSVL 8

RESULT 5
US-09-732-210-617
Sequence 617, Application US/09732210
Patent No. 6573361
GENERAL INFORMATION:
APPLICANT: Bunker, Greg J.
APPLICANT: Liang, Jihong
APPLICANT: Mitcanck, Cindy A.
APPLICANT: Seale, Jeffrey W.
APPLICANT: Wu, Yonnie S.
TITLE OF INVENTION: Anti-Fungal Proteins and Methods for Their Use
FILE REFERENCE: 38-21(15036)B
CURRENT APPLICATION NUMBER: US/09/732,210
```

```
CURRENT FILING DATE: 2000-12-07
PRIOR APPLICATION NUMBER: US 60/169,513
PRIOR FILING DATE: 1999-12-07
PRIOR APPLICATION NUMBER: US 60/169,340
PRIOR FILING DATE: 1999-12-07
NUMBER OF SEQ ID NOS: 1753
SEQ ID NO 617
LENGTH: 143
TYPE: PRT
ORGANISM: Methanococcus vannielii
US-09-732-210-617

Query Match      42.9%; Score 6; DB 4; Length 143;
Best Local Similarity 100.0%; Pred. No. 45;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MIRSK 6
DB 1 MIRSK 6

RESULT 6
US-09-543-681A-7075
Sequence 7075, Application US/09543681A
Patent No. 6605709
GENERAL INFORMATION:
APPLICANT: GARY BRETON
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
FILE REFERENCE: 2709.1002-001
CURRENT APPLICATION NUMBER: US/09/543, 681A
CURRENT FILING DATE: 2000-04-05
PRIOR APPLICATION NUMBER: US 60/128,706
PRIOR FILING DATE: 1999-04-09
NUMBER OF SEQ ID NOS: 8344
SEQ ID NO 7075
LENGTH: 183
TYPE: PRT
ORGANISM: Proteus mirabilis
US-09-543-681A-7075

Query Match      42.9%; Score 6; DB 4; Length 183;
Best Local Similarity 100.0%; Pred. No. 56;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 TSUUSF 13
DB 124 TSUUSF 129

RESULT 7
US-09-583-110-3590
Sequence 3590, Application US/09583110
Patent No. 6699703
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stramm et al.
TITLE OF INVENTION: Nucleic Acid and Amino Acid Sequences Relating to Streptococcus
FILE REFERENCE: PAT00-07A
CURRENT APPLICATION NUMBER: US/09/583, 110
CURRENT FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/107,433
PRIOR FILING DATE: 1998-06-30
PRIOR APPLICATION NUMBER: US 60/085,131
PRIOR FILING DATE: 1998-05-12
PRIOR APPLICATION NUMBER: US 60/051,553
PRIOR FILING DATE: 1997-07-02
NUMBER OF SEQ ID NOS: 5322
SEQ ID NO 3590
LENGTH: 238
TYPE: PRT
ORGANISM: Streptococcus pneumoniae
US-09-583-110-3590
```

Query Match 42.9%; Score 6; DB 4; Length 238;  
Best Local Similarity 100.0%; Pred. No. 70;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 KSKITS 9  
|||||  
Db 162 KSKITS 167

RESULT 8  
US-09-107-532A-3734  
Sequence 3734, Application US/09107532A  
Patent No. 6583275  
GENERAL INFORMATION:  
APPLICANT: Lynn A Doucette-Stamm and David Bush  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
NUMBER OF SEQUENCES: 7310  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
STREET: 100 Beaver Street  
CITY: Waltham  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02354  
COMPUTER READABLE FORM:  
MEDIUM TYPE: CD-ROM ISO9660  
COMPUTER: PC  
OPERATING SYSTEM: <Unknown>  
SOFTWARE: ASCII  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/107,532A  
FILING DATE: 30-Jun-1998  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/085,598  
FILING DATE: 14 May 1998  
APPLICATION NUMBER: 60/051571  
FILING DATE: July 2, 1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Atinello, Pamela Deneke  
REGISTRATION NUMBER: 40,489  
REFERENCE/DOCKET NUMBER: GTC-012  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (781)893-5007  
TELEFAX: (781)893-8277  
INFORMATION FOR SEQ ID NO: 3734:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 270 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHEICAL: YES  
ORIGINAL SOURCE:  
ORGANISM: Enterococcus faecium  
FEATURE:  
NAME/KEY: misc feature  
LOCATION: (B) LOCATION 1...270  
SEQUENCE DESCRIPTION: SEQ ID NO: 3734:  
US-09-107-532A-3734

Query Match 42.9%; Score 6; DB 4; Length 270;  
Best Local Similarity 100.0%; Pred. No. 77;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ITSVLS 12  
|||||  
Db 238 ITSVLS 243

RESULT 9  
US-09-248-796A-16584  
Sequence 16584, Application US/09248796A

Patent No. 6747137  
GENERAL INFORMATION:  
APPLICANT: Keith Weinstock et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS  
TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 107196.132  
CURRENT APPLICATION NUMBER: US/09/248,796A  
CURRENT FILING DATE: 1999-02-12  
PRIOR APPLICATION NUMBER: US 60/074,725  
PRIOR FILING DATE: 1998-02-13  
PRIOR APPLICATION NUMBER: US 60/096,409  
PRIOR FILING DATE: 1998-08-13  
NUMBER OF SEQ ID NOS: 28208  
SEQ ID NO 16584  
LENGTH: 275  
TYPE: PRT  
ORGANISM: Candida albicans  
US-09-248-796A-16584

Query Match 42.9%; Score 6; DB 4; Length 275;  
Best Local Similarity 100.0%; Pred. No. 79;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ITSVLS 12  
|||||  
Db 101 ITSVLS 106

RESULT 10  
US-09-489-039A-9786  
Sequence 9786, Application US/09489039A  
Patent No. 6610836  
GENERAL INFORMATION:  
APPLICANT: Gary Breton et. al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA  
FILE REFERENCE: 2709.2004001  
CURRENT APPLICATION NUMBER: US/09/489,039A  
CURRENT FILING DATE: 2000-01-27  
PRIOR APPLICATION NUMBER: US 60/117,747  
PRIOR FILING DATE: 1999-01-29  
NUMBER OF SEQ ID NOS: 14342  
SEQ ID NO 9786  
LENGTH: 311  
TYPE: PRT  
ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-9786

Query Match 42.9%; Score 6; DB 4; Length 311;  
Best Local Similarity 100.0%; Pred. No. 87;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ITSVLS 12  
|||||  
Db 261 ITSVLS 266

RESULT 11  
US-09-526-597D-34  
Sequence 34, Application US/09526597D  
Patent No. 6710227  
GENERAL INFORMATION:  
APPLICANT: De Veylder, Lieven  
APPLICANT: De Almeida, Janelle  
APPLICANT: Landrieu, Isabelle  
TITLE OF INVENTION: Cyclin-dependent kinase inhibitors and uses thereof  
FILE REFERENCE: 1187-2  
CURRENT APPLICATION NUMBER: US/09/526,597D  
CURRENT FILING DATE: 2000-03-16  
NUMBER OF SEQ ID NOS: 35  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 34  
LENGTH: 327



TYPE: PRT  
ORGANISM: Arabidopsis thaliana  
US-09-526-597D-34

Query Match 42.9%; Score 6; DB 4; Length 327;  
Best Local Similarity 100.0%; Pred. No. 91;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 IRSKI 7  
|||||  
Db 5 IRSKI 10

RESULT 12  
US-09-107-532A-5546  
Sequence 5546, Application US/09107532A  
Patent No. 6583275

GENERAL INFORMATION:

APPLICANT: Lynn A Doucette-Stamm and David Bush

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 7310

CORRESPONDENCE ADDRESS:

ADDRESSEE: GENOME THERAPEUTICS CORPORATION

STREET: 100 Beaver Street

CITY: Waltham

STATE: Massachusetts

COUNTRY: USA

ZIP: 02354

COMPUTER READABLE FORM:

MEDIUM TYPE: CD-ROM ISO9660

OPERATING SYSTEM: <Unknown>

SOFTWARE: ASCII

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/107,532A

FILING DATE: 30-Jun-1998

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/085,598

FILING DATE: 14 May 1998

APPLICATION NUMBER: 60/051571

FILING DATE: July 2, 1997

ATTORNEY/AGENT INFORMATION:

NAME: Ariniello, Pamela Deneke

REGISTRATION NUMBER: 40,489

REFERENCE/DOCKET NUMBER: GTC-012

TELECOMMUNICATION INFORMATION:

TELEPHONE: (781)893-5007

TELEFAX: (781)893-8277

INFORMATION FOR SEQ ID NO: 5546:

SEQUENCE CHARACTERISTICS:

LENGTH: 344 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHEICAL: YES

ORIGINAL SOURCE:

ORGANISM: Enterococcus faecium

FEATURE:

NAME/KEY: misc feature

LOCATION: (B) LOCATION 1..344

SEQUENCE DESCRIPTION: SEQ ID NO: 5546:

Query Match 42.9%; Score 6; DB 4; Length 344;  
Best Local Similarity 100.0%; Pred. No. 95;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 KSKITS 9  
|||||  
Db 163 KSKITS 168

RESULT 13  
US-08-700-359-20

Sequence 20, Application US/08700359

Patent No. 5766925

GENERAL INFORMATION:

APPLICANT: SUGIMOTO, MASAKAZU

APPLICANT: USUDA, YOSHIHIRO

APPLICANT: SUZUKI, TOMOKO

APPLICANT: TANAKA, AKIKO

APPLICANT: MATSUI, HIROSHI

TITLE OF INVENTION: METHOD OF PRODUCING L-LYSINE

NUMBER OF SEQUENCES: 22

CORRESPONDENCE ADDRESS:

ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MATER & NEUSTADT, P.C.

STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400

CITY: ARLINGTON

STATE: VA

COUNTRY: USA

ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: MS-DOS TEXT EDITOR

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/700,359

FILING DATE: 08-OCT-1996

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP-35019

FILING DATE: 04-MAR-1994

ATTORNEY/AGENT INFORMATION:

NAME: NORMAN F. OBLON

REGISTRATION NUMBER: 24,618

REFERENCE/DOCKET NUMBER: 10-819-0 PCT

TELECOMMUNICATION INFORMATION:

TELEPHONE: 703-413-3000

TELEFAX: 703-413-2220

INFORMATION FOR SEQ ID NO: 20:

SEQUENCE CHARACTERISTICS:

LENGTH: 433 amino acids

TYPE: amino acid

TOPOLOGY: linear

US-08-700-359-20

Query Match 42.9%; Score 6; DB 1; Length 433;  
Best Local Similarity 100.0%; Pred. No. 1.2e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 SKITSV 10  
|||||  
Db 363 SKITSV 368

RESULT 14  
US-09-019-095A-24

Sequence 24, Application US/09019095A

Patent No. 6287858

GENERAL INFORMATION:

APPLICANT: D'Andrea, Alan D.

APPLICANT: Zhu, Yuan

TITLE OF INVENTION: Deubiquitinating Enzymes That Regulate

CELL GROWTH

FILE REFERENCE: DFC1-435P2A2

CURRENT APPLICATION NUMBER: US/09/019,095A

PRIOR FILING DATE: 1998-02-05

PRIOR APPLICATION NUMBER: PCT/US96/12884

PRIOR FILING DATE: 1996-08-07

PRIOR APPLICATION NUMBER: US 60/002,066

PRIOR FILING DATE: 1995-08-09

PRIOR APPLICATION NUMBER: US 60/019,787

PRIOR FILING DATE: 1996-06-14

NUMBER OF SEQ ID NOS: 51

```
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 24
; LENGTH: 494
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-019-095A-24
```

```
Query Match          42.9%; Score 6; DB 3; Length 494;
Best Local Similarity 100.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 7 ITSVLS 12
    |||||
Db 324 ITSVLS 329
```

```
RESULT 15
US-08-945-771-2
; Sequence 2, Application US/08945771
; Patent No. 6465623
; GENERAL INFORMATION:
; APPLICANT: Daly, Roger J
; APPLICANT: Sutherland, Robert L
; TITLE OF INVENTION: GDU, A novel signalling protein
; FILE REFERENCE: 273402001700
; CURRENT APPLICATION NUMBER: US/08/945, 771
; CURRENT FILING DATE: 1998-04-22
; EARLIER APPLICATION NUMBER: PCT/US96/00258
; EARLIER FILING DATE: 1996-MAY-02
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 540
; TYPE: PRT
; ORGANISM: Homo sapiens
US-08-945-771-2
```

```
Query Match          42.9%; Score 6; DB 4; Length 540;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 7 ITSVLS 12
    |||||
Db 90 ITSVLS 95
```

```
RESULT 16
US-09-371-338-7
; Sequence 7, Application US/09371338
; Patent No. 6613959
; GENERAL INFORMATION:
; APPLICANT: Sheen, Jen
; APPLICANT: Kovtun, Yelena V.
; APPLICANT: Chiu, Wan-Ling
; TITLE OF INVENTION: TRANSGENIC PLANTS EXPRESSING A MAPKKK
; FILE REFERENCE: 00786/366002
; CURRENT APPLICATION NUMBER: US/09/371,338
; CURRENT FILING DATE: 1999-08-10
; PRIOR APPLICATION NUMBER: 60/095,938
; PRIOR FILING DATE: 1998-08-10
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 661
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-371-338-7
```

```
Query Match          42.9%; Score 6; DB 4; Length 661;
Best Local Similarity 100.0%; Pred. No. 1.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 2 IRKSKI 7
    |||||
Db 36 IRKSKI 41
```

```
RESULT 17
US-09-134-001C-3586
; Sequence 3586, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3586
; LENGTH: 739
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3586
```

```
Query Match          42.9%; Score 6; DB 3; Length 739;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 8 TSVLSF 13
    |||||
Db 446 TSVLSF 451
```

```
RESULT 18
US-09-543-681A-6407
; Sequence 6407, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709,1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 6407
; LENGTH: 967
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-6407
```

```
Query Match          42.9%; Score 6; DB 4; Length 967;
Best Local Similarity 100.0%; Pred. No. 2.3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 7 ITSVLS 12
    |||||
Db 781 ITSVLS 786
```

```
RESULT 19
US-09-177-249-124
; Sequence 124, Application US/09177249
; Patent No. 6223064
; GENERAL INFORMATION:
; APPLICANT: Fischer, Robert L.
; APPLICANT: Ohad, Nir
; APPLICANT: Kiyosue, Tomohiro
```

APPLICANT: Yadegari, Ramlin  
APPLICANT: Margossian, Linda  
APPLICANT: Harada, John  
APPLICANT: Goldberg, Robert B.  
APPLICANT: The Regents of the University of California  
TITLE OF INVENTION: Nucleic Acids That Control Seed and Fruit  
FILE REFERENCE: 023070-086120US  
CURRENT APPLICATION NUMBER: US/09/177,249  
CURRENT FILING DATE: 1998-10-22  
EARLIER APPLICATION NUMBER: US 09/071,838  
EARLIER FILING DATE: 1998-05-01  
NUMBER OF SEQ ID NOS: 324  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 124  
LENGTH: 12  
TYPE: PRT  
ORGANISM: Arabidopsis sp.  
US-09-177-249-124

Query Match 35.7%; Score 5; DB 3; Length 12;  
Best Local Similarity 100.0%; Pred. No. 55;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ITSVL 11  
Db 5 ITSVL 9

RESULT 20  
US-09-522-666-30  
Sequence 30, Application US/09522666  
Patent No. 633167  
GENERAL INFORMATION:  
APPLICANT: Shuey, David  
APPLICANT: Quinet, Elaine  
TITLE OF INVENTION: Methods and Reagents for Identifying Inhibitors of  
TITLE OF INVENTION: Protocols of Membrane-Associated Proteins  
FILE REFERENCE: 6-00  
CURRENT APPLICATION NUMBER: US/09/522,666  
CURRENT FILING DATE: 2000-03-10  
NUMBER OF SEQ ID NOS: 32  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 30  
LENGTH: 13  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: amino acid  
OTHER INFORMATION: sequence encompassing Site-1 protease cleavage  
OTHER INFORMATION: site within SREPB-2  
US-09-522-666-30

Query Match 35.7%; Score 5; DB 3; Length 13;  
Best Local Similarity 100.0%; Pred. No. 59;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 SVLSF 13  
Db 8 SVLSF 12

RESULT 21  
US-08-268-251-5  
Sequence 5, Application US/08268251  
Patent No. 5585475  
GENERAL INFORMATION:  
APPLICANT: Jamieson, Gordon A  
APPLICANT: Dedman, John R  
APPLICANT: Kaetzel, Marcia A  
TITLE OF INVENTION: Calmodulin-Binding Peptides  
NUMBER OF SEQUENCES: 59  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Chiron Corporation  
STREET: 4560 Horton Street  
CITY: Emeryville  
STATE: CA  
COUNTRY: USA  
ZIP: 94608  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WordPerfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/268,251  
FILING DATE:  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/831,219  
FILING DATE: 06-FEB-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Green, Grant D  
REGISTRATION NUMBER: 31,259  
REFERENCE/DOCKET NUMBER: 272,001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (510) 601-2706  
TELEFAX: (510) 655-3542  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 15 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-268-251-5

Query Match 35.7%; Score 5; DB 1; Length 15;  
Best Local Similarity 100.0%; Pred. No. 67;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 TSVAL 12  
Db 9 TSVAL 13

RESULT 22  
PCT-US93-01112-5  
Sequence 5, Application PC/TUS9301112  
GENERAL INFORMATION:  
APPLICANT: Jamieson, Gordon A  
APPLICANT: Dedman, John R  
APPLICANT: Kaetzel, Marcia A  
TITLE OF INVENTION: Calmodulin-Binding Peptides  
NUMBER OF SEQUENCES: 59  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Chiron Corporation  
STREET: 4560 Horton Street  
CITY: Emeryville  
STATE: CA  
COUNTRY: USA  
ZIP: 94608  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WordPerfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/01112  
FILING DATE: 19930208  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/831,219  
FILING DATE: 06-FEB-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Green, Grant D

REGISTRATION NUMBER: 31,259  
REFERENCE/DOCKET NUMBER: 272,001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (510) 601-2706  
TELEFAX: (510) 655-3542  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 15 amino acids  
TYPE: AMINO ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
PCT-US93-01112-5

Query Match 35.7%; Score 5; DB 5; Length 15;  
Best Local Similarity 100.0%; Pred. No. 67;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 TSVALS 12  
Db 9 TSVALS 13

RESULT 23  
US-09-360-237-55  
Sequence 55, Application US/09360237  
Patent No. 6322962  
GENERAL INFORMATION:  
APPLICANT: BROWN, MICHAEL S.  
APPLICANT: CHENG, DONG  
APPLICANT: ESPENSHADE, PETER J.  
APPLICANT: GOLDSTEIN, JOSEPH L.  
APPLICANT: RAMSON, ROBERT B.  
APPLICANT: SAKAI, JURO  
TITLE OF INVENTION: STEROL-REGULATED SITE-1 PROTEASE AND ASSAYS OF  
TITLE OF INVENTION: MODULATORS THEREOF  
FILE REFERENCE: UTXD:567  
CURRENT APPLICATION NUMBER: US/09/360,237  
CURRENT FILING DATE: 1999-07-23  
EARLIER APPLICATION NUMBER: 60/096,571  
EARLIER FILING DATE: 1998-08-14  
NUMBER OF SEQ ID NOS: 60  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 55  
LENGTH: 16  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: SYNTHETIC  
US-09-360-237-55

Query Match 35.7%; Score 5; DB 3; Length 16;  
Best Local Similarity 100.0%; Pred. No. 70;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 SVLSF 13  
Db 7 SVLSF 11

RESULT 24  
US-09-556-877-251  
Sequence 251, Application US/09556877  
Patent No. 6432916  
GENERAL INFORMATION:  
APPLICANT: Probst, Peter  
APPLICANT: Bhatia, Ajay  
APPLICANT: Skeiky, Yasir  
APPLICANT: Fling, Steve  
APPLICANT: Maisonneuve, Jeff  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND  
TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION

FILE REFERENCE: 210121.469C5  
CURRENT APPLICATION NUMBER: US/09/556,877  
CURRENT FILING DATE: 2000-04-19  
NUMBER OF SEQ ID NOS: 305  
SOFTWARE: FastSeq for Windows Version 3.0/4.0  
SEQ ID NO 251  
LENGTH: 16  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Made in a lab  
US-09-556-877-251

Query Match 35.7%; Score 5; DB 4; Length 16;  
Best Local Similarity 100.0%; Pred. No. 70;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 KSKIT 8  
Db 3 KSKIT 7

RESULT 25  
US-09-620-412C-251  
Sequence 251, Application US/09620412C  
Patent No. 6448234  
GENERAL INFORMATION:  
APPLICANT: Steven P. Fling  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND  
TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION  
FILE REFERENCE: 210121.469C7  
CURRENT APPLICATION NUMBER: US/09/620,412C  
CURRENT FILING DATE: 2000-07-20  
NUMBER OF SEQ ID NOS: 363  
SOFTWARE: FastSeq for Windows Version 3.0/4.0  
SEQ ID NO 251  
LENGTH: 16  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Made in a lab  
US-09-620-412C-251

Query Match 35.7%; Score 5; DB 4; Length 16;  
Best Local Similarity 100.0%; Pred. No. 70;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 KSKIT 8  
Db 3 KSKIT 7

RESULT 26  
US-09-598-419-251  
Sequence 251, Application US/09598419  
Patent No. 6565856  
GENERAL INFORMATION:  
APPLICANT: Skeiky, Yasir A.W.  
APPLICANT: Scholler, John  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR TREATMENT AND  
TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION  
FILE REFERENCE: 210121.469C6  
CURRENT APPLICATION NUMBER: US/09/598,419  
CURRENT FILING DATE: 2000-06-20  
NUMBER OF SEQ ID NOS: 357  
SOFTWARE: FastSeq for Windows Version 3.0/4.0  
SEQ ID NO 251  
LENGTH: 16  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Made in a lab  
US-09-598-419-251

Query Match 35.7%; Score 5; DB 4; Length 16;  
Best Local Similarity 100.0%; Pred. No. 70;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 KSKIT 8  
|||  
Db 3 KSKIT 7

## RESULT 27

US-09-680-571A-83  
; Sequence 83, Application US/09680571A  
; Patent No. 6649593  
; GENERAL INFORMATION:  
; APPLICANT: Jaen, Juan C.  
; APPLICANT: Li, Leping  
; APPLICANT: Brown, Michael S.  
; APPLICANT: Goldstein, Joseph L.  
; APPLICANT: Cheng, Dong  
; APPLICANT: Tularik Inc.  
; APPLICANT: Board of Regents, The University of Texas System  
; APPLICANT: Southwestern Medical Center  
; TITLE OF INVENTION: Modulators of SREBP Processing  
; FILE REFERENCE: 018781-002810US  
; CURRENT APPLICATION NUMBER: US/09/680,571A  
; PRIOR FILING DATE: 2000-10-06  
; PRIOR APPLICATION NUMBER: US 60/159,236  
; PRIOR FILING DATE: 1999-10-13  
; NUMBER OF SEQ ID NOS: 107  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 83  
; LENGTH: 16  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: typical Site-1  
; OTHER INFORMATION: protease (S1P) peptide cleavage substrate  
US-09-680-571A-83

Query Match 35.7%; Score 5; DB 4; Length 16;  
Best Local Similarity 100.0%; Pred. No. 70;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 SVLSF 13  
|||  
Db 7 SVLSF 11

## RESULT 28

US-09-680-571A-97  
; Sequence 97, Application US/09680571A  
; Patent No. 6649593  
; GENERAL INFORMATION:  
; APPLICANT: Jaen, Juan C.  
; APPLICANT: Li, Leping  
; APPLICANT: Brown, Michael S.  
; APPLICANT: Goldstein, Joseph L.  
; APPLICANT: Cheng, Dong  
; APPLICANT: Tularik Inc.  
; APPLICANT: Board of Regents, The University of Texas System  
; APPLICANT: Southwestern Medical Center  
; TITLE OF INVENTION: Modulators of SREBP Processing  
; FILE REFERENCE: 018781-002810US  
; CURRENT APPLICATION NUMBER: US/09/680,571A  
; PRIOR FILING DATE: 2000-10-06  
; PRIOR APPLICATION NUMBER: US 60/159,236  
; PRIOR FILING DATE: 1999-10-13  
; NUMBER OF SEQ ID NOS: 107  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 97  
; LENGTH: 16  
; TYPE: PRT

ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthesized  
; OTHER INFORMATION: peptide  
US-09-680-571A-97

Query Match 35.7%; Score 5; DB 4; Length 16;  
Best Local Similarity 100.0%; Pred. No. 70;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 9 SVLSF 13  
|||  
Db 7 SVLSF 11

## RESULT 29

US-09-556-877-248  
; Sequence 248, Application US/09556877  
; Patent No. 6432916  
; GENERAL INFORMATION:  
; APPLICANT: Probst, Peter  
; APPLICANT: Bhacia, Ajay  
; APPLICANT: Skeiky, Yasir  
; APPLICANT: Fling, Steve  
; APPLICANT: Maisonneuve, Jeff  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND  
; TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION  
; FILE REFERENCE: 210121.469C5  
; CURRENT APPLICATION NUMBER: US/09/556,877  
; CURRENT FILING DATE: 2000-04-19  
; NUMBER OF SEQ ID NOS: 305  
; SOFTWARE: FastSeq for Windows Version 3.0/4.0  
; SEQ ID NO 248  
; LENGTH: 20  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Made in a lab  
US-09-556-877-248

Query Match 35.7%; Score 5; DB 4; Length 20;  
Best Local Similarity 100.0%; Pred. No. 85;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 KSKIT 8  
|||  
Db 12 KSKIT 16

## RESULT 30

US-09-556-877-249  
; Sequence 249, Application US/09556877  
; Patent No. 6432916  
; GENERAL INFORMATION:  
; APPLICANT: Probst, Peter  
; APPLICANT: Bhacia, Ajay  
; APPLICANT: Skeiky, Yasir  
; APPLICANT: Fling, Steve  
; APPLICANT: Maisonneuve, Jeff  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND  
; TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION  
; FILE REFERENCE: 210121.469C5  
; CURRENT APPLICATION NUMBER: US/09/556,877  
; CURRENT FILING DATE: 2000-04-19  
; NUMBER OF SEQ ID NOS: 305  
; SOFTWARE: FastSeq for Windows Version 3.0/4.0  
; SEQ ID NO 249  
; LENGTH: 20  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Made in a lab  
US-09-556-877-249

Query Match 35.7%; Score 5; DB 4; Length 20;  
Best Local Similarity 100.0%; Pred. No. 85;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 KSKIT 8  
      |||||  
Db 7 KSKIT 11

RESULT 31  
US-09-556-877-250  
; Sequence 250, Application US/09556877  
; Patent No. 6432916  
; GENERAL INFORMATION:  
; APPLICANT: Probst, Peter  
; APPLICANT: Bhatia, Ajay  
; APPLICANT: Skeiky, Yasir  
; APPLICANT: Fling, Steve  
; APPLICANT: Maisonneuve, Jeff  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND  
; TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION  
; FILE REFERENCE: 210121.469C5  
; CURRENT APPLICATION NUMBER: US/09/556, 877  
; CURRENT FILING DATE: 2000-04-19  
; NUMBER OF SEQ ID NOS: 305  
; SOFTWARE: FastSeq for Windows Version 3.0/4.0  
; SEQ ID NO 250  
; LENGTH: 20  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Made in a lab  
US-09-556-877-250

Query Match 35.7%; Score 5; DB 4; Length 20;  
Best Local Similarity 100.0%; Pred. No. 85;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 KSKIT 8  
      |||||  
Db 3 KSKIT 7

RESULT 32  
US-09-620-412C-248  
; Sequence 248, Application US/09620412C  
; Patent No. 6448234  
; GENERAL INFORMATION:  
; APPLICANT: Steven P. Fling  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND  
; TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION  
; FILE REFERENCE: 210121.469C7  
; CURRENT APPLICATION NUMBER: US/09/620, 412C  
; CURRENT FILING DATE: 2000-07-20  
; NUMBER OF SEQ ID NOS: 363  
; SOFTWARE: FastSeq for Windows Version 3.0/4.0  
; SEQ ID NO 248  
; LENGTH: 20  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Made in a lab  
US-09-620-412C-248

Query Match 35.7%; Score 5; DB 4; Length 20;  
Best Local Similarity 100.0%; Pred. No. 85;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 KSKIT 8  
      |||||  
Db 12 KSKIT 16

RESULT 33  
US-09-620-412C-249  
; Sequence 249, Application US/09620412C  
; Patent No. 6448234  
; GENERAL INFORMATION:  
; APPLICANT: Steven P. Fling  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND  
; TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION  
; FILE REFERENCE: 210121.469C7  
; CURRENT APPLICATION NUMBER: US/09/620, 412C  
; CURRENT FILING DATE: 2000-07-20  
; NUMBER OF SEQ ID NOS: 363  
; SOFTWARE: FastSeq for Windows Version 3.0/4.0  
; SEQ ID NO 249  
; LENGTH: 20  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Made in a lab  
US-09-620-412C-249

Query Match 35.7%; Score 5; DB 4; Length 20;  
Best Local Similarity 100.0%; Pred. No. 85;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 KSKIT 8  
      |||||  
Db 7 KSKIT 11

RESULT 34  
US-09-620-412C-250  
; Sequence 250, Application US/09620412C  
; Patent No. 6448234  
; GENERAL INFORMATION:  
; APPLICANT: Steven P. Fling  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND  
; TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION  
; FILE REFERENCE: 210121.469C7  
; CURRENT APPLICATION NUMBER: US/09/620, 412C  
; CURRENT FILING DATE: 2000-07-20  
; NUMBER OF SEQ ID NOS: 363  
; SOFTWARE: FastSeq for Windows Version 3.0/4.0  
; SEQ ID NO 250  
; LENGTH: 20  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Made in a lab  
US-09-620-412C-250

Query Match 35.7%; Score 5; DB 4; Length 20;  
Best Local Similarity 100.0%; Pred. No. 85;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 KSKIT 8  
      |||||  
Db 3 KSKIT 7

RESULT 35  
US-09-598-419-248  
; Sequence 248, Application US/09598419  
; Patent No. 6565856  
; GENERAL INFORMATION:  
; APPLICANT: Skeiky, Yasir A.W.  
; APPLICANT: Scholler, John  
; TITLE OF INVENTION: COMPOUNDS AND METHODS FOR TREATMENT AND  
; TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION  
; FILE REFERENCE: 210121.469C6  
; CURRENT APPLICATION NUMBER: US/09/598, 419  
; CURRENT FILING DATE: 2000-06-20

NUMBER OF SEQ ID NOS: 357  
SOFTWARE: FastSeq for Windows Version 3.0/4.0  
SEQ ID NO 248  
LENGTH: 20  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Made in a lab  
US-09-598-419-248

Query Match 35.7%; Score 5; DB 4; Length 20;  
Best Local Similarity 100.0%; Pred. No. 85;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 KSKIT 8  
Db 12 KSKIT 16

RESULT 36  
US-09-598-419-249  
Sequence 249, Application US/09598419  
Patent No. 6565856  
GENERAL INFORMATION:  
APPLICANT: Skeiky, Yasir A.W.  
APPLICANT: Scholler, John  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR TREATMENT AND  
FILE REFERENCE: 210121.469C6  
CURRENT APPLICATION NUMBER: US/09/598,419  
CURRENT FILING DATE: 2000-06-20  
NUMBER OF SEQ ID NOS: 357  
SOFTWARE: FastSeq for Windows Version 3.0/4.0  
SEQ ID NO 249  
LENGTH: 20  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Made in a lab  
US-09-598-419-249

Query Match 35.7%; Score 5; DB 4; Length 20;  
Best Local Similarity 100.0%; Pred. No. 85;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 KSKIT 8  
Db 7 KSKIT 11

RESULT 37  
US-09-598-419-250  
Sequence 250, Application US/09598419  
Patent No. 6565856  
GENERAL INFORMATION:  
APPLICANT: Skeiky, Yasir A.W.  
APPLICANT: Scholler, John  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR TREATMENT AND  
FILE REFERENCE: 210121.469C6  
CURRENT APPLICATION NUMBER: US/09/598,419  
CURRENT FILING DATE: 2000-06-20  
NUMBER OF SEQ ID NOS: 357  
SOFTWARE: FastSeq for Windows Version 3.0/4.0  
SEQ ID NO 250  
LENGTH: 20  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Made in a lab  
US-09-598-419-250

Query Match 35.7%; Score 5; DB 4; Length 20;

Best Local Similarity 100.0%; Pred. No. 85;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 KSKIT 8  
Db 3 KSKIT 7

RESULT 38  
US-09-834-759-545  
Sequence 545, Application US/09834759  
Patent No. 6680197  
GENERAL INFORMATION:  
APPLICANT: Jiang, Yugu  
APPLICANT: Dillon, Devin C.  
APPLICANT: Mitcham, Jennifer L.  
APPLICANT: Xu, Jiangchun  
APPLICANT: Harlocker, Susan L.  
APPLICANT: Hepler, William T.  
APPLICANT: Henderson, Robert A.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND  
FILE REFERENCE: 210121.470C9  
CURRENT APPLICATION NUMBER: US/09/834,759  
CURRENT FILING DATE: 2001-04-13  
NUMBER OF SEQ ID NOS: 547  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 545  
LENGTH: 20  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-834-759-545

Query Match 35.7%; Score 5; DB 4; Length 20;  
Best Local Similarity 100.0%; Pred. No. 85;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 KSKIT 8  
Db 3 KSKIT 7

RESULT 39  
US-08-268-251-35  
Sequence 35, Application US/08268251  
Patent No. 3585475  
GENERAL INFORMATION:  
APPLICANT: Jamieson, Gordon A  
APPLICANT: Dedman, John R  
APPLICANT: Kaetzel, Marcia A  
TITLE OF INVENTION: Calmodulin-Binding Peptides  
NUMBER OF SEQUENCES: 59  
CORRESPONDENCE ADDRESS:  
ADDRESS: Chiron Corporation  
STREET: 4560 Horton Street  
CITY: Emeryville  
STATE: CA  
COUNTRY: USA  
ZIP: 94608  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WordPerfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/268,251  
FILING DATE:  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/831,219  
FILING DATE: 06-FEB-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Green, Grant D

REGISTRATION NUMBER: 31,259  
REFERENCE/DOCKET NUMBER: 272.001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (510) 601-2706  
TELEFAX: (510) 655-3542  
INFORMATION FOR SEQ ID NO: 35:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 23 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-268-251-35

Query Match 35.7%; Score 5; DB 1; Length 23;  
Best Local Similarity 100.0%; Pred. No. 96;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 TSVLS 12  
|||  
Db 11 TSVLS 15

RESULT 40  
PCT-US93-01112-35  
Sequence 35, Application PC/TUS9301112  
GENERAL INFORMATION:  
APPLICANT: Jamieson, Gordon A  
APPLICANT: Dedman, John R  
APPLICANT: Kaetzl, Marcia A  
TITLE OF INVENTION: Calmodulin-Binding Peptides  
NUMBER OF SEQUENCES: 59  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Chiron Corporation  
STREET: 4560 Horton Street  
CITY: Emeryville  
STATE: CA  
COUNTRY: USA  
ZIP: 94608  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WordPerfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/01112  
FILING DATE: 19930208  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/831,219  
FILING DATE: 06-FEB-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Green, Grant D  
REGISTRATION NUMBER: 31,259  
REFERENCE/DOCKET NUMBER: 272.001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (510) 601-2706  
TELEFAX: (510) 655-3542  
INFORMATION FOR SEQ ID NO: 35:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 23 amino acids  
TYPE: AMINO ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
PCT-US93-01112-35

Query Match 35.7%; Score 5; DB 5; Length 23;  
Best Local Similarity 100.0%; Pred. No. 96;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 TSVLS 12  
|||  
Db 11 TSVLS 15

Db 11 TSVLS 15

Search completed: December 30, 2004, 15:37:58  
Job time: 42 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: December 30, 2004, 15:34:09 ; Search time 144 Seconds  
(without alignments)  
34.974 Million cell updates/sec

Title: US-10-719-385-20  
Perfect score: 14  
Sequence: 1 MIRSKITSVLSFC 14

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 1599051 seqs, 359727711 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1599051

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: listing first 150 summaries

## Database :

Published Applications MA:\*

- 1: /cgn2\_6/ptodata/1/pubppaa/US07\_PUBCOMB.pep:\*
- 2: /cgn2\_6/ptodata/1/pubppaa/PCT\_NEW\_PUB.pep:\*
- 3: /cgn2\_6/ptodata/1/pubppaa/US06\_NEW\_PUB.pep:\*
- 4: /cgn2\_6/ptodata/1/pubppaa/US06\_PUBCOMB.pep:\*
- 5: /cgn2\_6/ptodata/1/pubppaa/US07\_NEW\_PUB.pep:\*
- 6: /cgn2\_6/ptodata/1/pubppaa/PCTUS\_PUBCOMB.pep:\*
- 7: /cgn2\_6/ptodata/1/pubppaa/US08\_NEW\_PUB.pep:\*
- 8: /cgn2\_6/ptodata/1/pubppaa/US08\_PUBCOMB.pep:\*
- 9: /cgn2\_6/ptodata/1/pubppaa/US09\_PUBCOMB.pep:\*
- 10: /cgn2\_6/ptodata/1/pubppaa/US09\_PUBCOMB.pep:\*
- 11: /cgn2\_6/ptodata/1/pubppaa/US09C\_PUBCOMB.pep:\*
- 12: /cgn2\_6/ptodata/1/pubppaa/US09\_NEW\_PUB.pep:\*
- 13: /cgn2\_6/ptodata/1/pubppaa/US10\_PUBCOMB.pep:\*
- 14: /cgn2\_6/ptodata/1/pubppaa/US10\_PUBCOMB.pep:\*
- 15: /cgn2\_6/ptodata/1/pubppaa/US10C\_PUBCOMB.pep:\*
- 16: /cgn2\_6/ptodata/1/pubppaa/US10D\_PUBCOMB.pep:\*
- 17: /cgn2\_6/ptodata/1/pubppaa/US10\_NEW\_PUB.pep:\*
- 18: /cgn2\_6/ptodata/1/pubppaa/US11\_NEW\_PUB.pep:\*
- 19: /cgn2\_6/ptodata/1/pubppaa/US60\_NEW\_PUB.pep:\*
- 20: /cgn2\_6/ptodata/1/pubppaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	14	100.0	14 17 US-10-719-385-20	Sequence 20, Appl
2	14	100.0	1752 17 US-10-719-385-5	Sequence 5, Appl1
3	14	100.0	1753 17 US-10-719-385-2	Sequence 2, Appl1
4	14	100.0	1753 17 US-10-719-385-3	Sequence 3, Appl1
5	14	100.0	1753 17 US-10-719-385-4	Sequence 4, Appl1
6	14	100.0	1753 17 US-10-719-385-6	Sequence 6, Appl1
7	14	100.0	1753 17 US-10-719-385-7	Sequence 7, Appl1
8	14	100.0	1753 17 US-10-719-385-8	Sequence 8, Appl1
9	14	100.0	1753 17 US-10-719-385-9	Sequence 9, Appl1
10	14	100.0	1753 17 US-10-719-385-10	Sequence 10, Appl
11	14	100.0	1753 17 US-10-719-385-11	Sequence 11, Appl
12	14	100.0	1753 17 US-10-719-385-12	Sequence 12, Appl
13	14	100.0	1753 17 US-10-719-385-13	Sequence 13, Appl

14	14	100.0	1753 17 US-10-719-385-14	Sequence 14, Appl
15	14	100.0	1753 17 US-10-719-385-15	Sequence 15, Appl
16	14	100.0	1753 17 US-10-719-385-16	Sequence 16, Appl
17	14	100.0	1753 17 US-10-719-385-17	Sequence 17, Appl
18	14	100.0	1753 17 US-10-719-385-18	Sequence 18, Appl
19	14	100.0	1753 17 US-10-719-385-19	Sequence 19, Appl
20	14	100.0	1753 17 US-10-719-385-20	Sequence 20, Appl
21	7	50.0	285 15 US-10-424-599-148727	Sequence 148727, A
22	7	50.0	290 15 US-10-425-114-47775	Sequence 47775, A
23	7	50.0	363 15 US-10-424-599-158692	Sequence 158692, A
24	7	50.0	431 9 US-09-815-242-10735	Sequence 10735, A
25	7	50.0	15 US-10-282-122A-57038	Sequence 57038, A
26	7	50.0	1019 9 US-09-738-626-6499	Sequence 6499, Ap
27	7	50.0	1019 9 US-10-380-055-12	Sequence 12, Appl
28	7	50.0	1065 15 US-10-424-599-148728	Sequence 148728, A
29	6	42.9	12 14 US-10-075-869-10	Sequence 10, Appl
30	6	42.9	12 14 US-10-366-493-10	Sequence 10, Appl
31	6	42.9	17 14 US-10-029-347-10	Sequence 10, Appl
32	6	42.9	48 16 US-10-437-963-111764	Sequence 111764, A
33	6	42.9	53 15 US-10-424-599-267615	Sequence 267615, A
34	6	42.9	56 17 US-10-425-115-185349	Sequence 185349, A
35	6	42.9	60 16 US-10-437-963-107926	Sequence 107926, A
36	6	42.9	69 15 US-10-424-599-284793	Sequence 284793, A
37	6	42.9	82 15 US-10-424-599-143794	Sequence 143794, A
38	6	42.9	100 15 US-10-424-599-217068	Sequence 217068, A
39	6	42.9	100 15 US-10-425-115-307173	Sequence 307173, A
40	6	42.9	111 15 US-10-425-114-54103	Sequence 54103, A
41	6	42.9	130 15 US-10-282-122A-45474	Sequence 45474, A
42	6	42.9	151 15 US-10-425-114-44778	Sequence 44778, A
43	6	42.9	178 16 US-10-767-701-38935	Sequence 38935, A
44	6	42.9	184 15 US-10-370-838-9	Sequence 9, Appl1
45	6	42.9	184 15 US-10-789-433-12	Sequence 12, Appl
46	6	42.9	186 15 US-10-370-838-8	Sequence 8, Appl1
47	6	42.9	186 17 US-10-789-433-14	Sequence 14, Appl
48	6	42.9	187 17 US-10-789-433-9	Sequence 9, Appl1
49	6	42.9	187 17 US-10-789-433-10	Sequence 10, Appl
50	6	42.9	198 16 US-10-437-963-129082	Sequence 129082, A
51	6	42.9	258 17 US-10-739-930-9132	Sequence 9132, Ap
52	6	42.9	289 15 US-10-451-139-17	Sequence 17, Appl
53	6	42.9	302 17 US-10-149-826-11	Sequence 11, Appl
54	6	42.9	304 17 US-10-425-115-270799	Sequence 270799, A
55	6	42.9	311 9 US-09-886-055-15	Sequence 15, Appl
56	6	42.9	311 15 US-09-804-291-15	Sequence 15, Appl
57	6	42.9	311 15 US-10-343-650A-402	Sequence 402, Ap
58	6	42.9	314 14 US-10-025-806-18	Sequence 18, Appl
59	6	42.9	317 15 US-09-965-621-60	Sequence 60, Appl
60	6	42.9	317 15 US-10-407-866-60	Sequence 60, Appl
61	6	42.9	317 16 US-10-781-294-60	Sequence 60, Appl
62	6	42.9	321 15 US-10-407-866-107	Sequence 107, Ap
63	6	42.9	321 15 US-10-264-237-2549	Sequence 2549, Ap
64	6	42.9	326 14 US-10-160-764-5	Sequence 5, Appl1
65	6	42.9	326 15 US-10-229-541A-11	Sequence 11, Appl
66	6	42.9	326 17 US-10-773-529-20	Sequence 20, Appl
67	6	42.9	327 15 US-10-688-291-34	Sequence 34, Appl
68	6	42.9	360 10 US-09-372-348-9	Sequence 9, Appl1
69	6	42.9	374 15 US-10-297-021-4	Sequence 4, Appl1
70	6	42.9	384 17 US-10-128-558-205	Sequence 205, Ap
71	6	42.9	387 17 US-10-408-865A-2232	Sequence 2232, Ap
72	6	42.9	403 16 US-10-469-864-2	Sequence 2, Appl1
73	6	42.9	404 16 US-10-474-776-724	Sequence 724, Ap
74	6	42.9	431 15 US-10-282-122A-57773	Sequence 57773, A
75	6	42.9	433 14 US-10-369-493-17137	Sequence 17137, A
76	6	42.9	433 14 US-10-369-493-23280	Sequence 23280, A
77	6	42.9	458 14 US-10-369-493-3910	Sequence 3910, Ap
78	6	42.9	497 14 US-10-371-905A-28	Sequence 28, Appl
79	6	42.9	530 14 US-10-107-695B-2	Sequence 2, Appl1
80	6	42.9	530 14 US-10-371-905A-6	Sequence 6, Appl1
81	6	42.9	530 14 US-10-371-905A-12	Sequence 12, Appl
82	6	42.9	530 14 US-10-371-905A-16	Sequence 16, Appl
83	6	42.9	530 14 US-10-371-905A-18	Sequence 18, Appl
84	6	42.9	530 14 US-10-371-905A-30	Sequence 30, Appl
85	6	42.9	530 14 US-10-371-905A-32	Sequence 32, Appl

87	6	42.9	530	14	US-10-371-905A-34	Sequence 34, Appl
88	6	42.9 <td>530</td> <td>14<th>US-10-371-905A-36</th><th>Sequence 36, Appl</th></td>	530	14 <th>US-10-371-905A-36</th> <th>Sequence 36, Appl</th>	US-10-371-905A-36	Sequence 36, Appl
89	6	42.9 <td>540</td> <td>14<th>US-08-945-771-2</th><th>Sequence 2, Appl1</th></td>	540	14 <th>US-08-945-771-2</th> <th>Sequence 2, Appl1</th>	US-08-945-771-2	Sequence 2, Appl1
90	6	42.9 <td>540</td> <td>14<th>US-10-242-332-2</th><th>Sequence 2, Appl1</th></td>	540	14 <th>US-10-242-332-2</th> <th>Sequence 2, Appl1</th>	US-10-242-332-2	Sequence 2, Appl1
91	6	42.9 <td>540</td> <td>14<th>US-10-323-001-2</th><th>Sequence 2, Appl1</th></td>	540	14 <th>US-10-323-001-2</th> <th>Sequence 2, Appl1</th>	US-10-323-001-2	Sequence 2, Appl1
92	6	42.9 <td>558</td> <td>15<th>US-10-282-122A-53651</th><th>Sequence 53651, A</th></td>	558	15 <th>US-10-282-122A-53651</th> <th>Sequence 53651, A</th>	US-10-282-122A-53651	Sequence 53651, A
93	6	42.9 <td>565</td> <td>14<th>US-10-371-905A-20</th><th>Sequence 20, Appl</th></td>	565	14 <th>US-10-371-905A-20</th> <th>Sequence 20, Appl</th>	US-10-371-905A-20	Sequence 20, Appl
94	6	42.9 <td>574</td> <td>14<th>US-10-371-905A-14</th><th>Sequence 14, Appl</th></td>	574	14 <th>US-10-371-905A-14</th> <th>Sequence 14, Appl</th>	US-10-371-905A-14	Sequence 14, Appl
95	6	42.9 <td>578</td> <td>15<th>US-10-428-275-14</th><th>Sequence 14, Appl</th></td>	578	15 <th>US-10-428-275-14</th> <th>Sequence 14, Appl</th>	US-10-428-275-14	Sequence 14, Appl
96	6	42.9 <td>584</td> <td>14<th>US-10-029-386-32446</th><th>Sequence 32446, A</th></td>	584	14 <th>US-10-029-386-32446</th> <th>Sequence 32446, A</th>	US-10-029-386-32446	Sequence 32446, A
97	6	42.9 <td>597</td> <td>15<th>US-10-112-944-770</th><th>Sequence 770, App</th></td>	597	15 <th>US-10-112-944-770</th> <th>Sequence 770, App</th>	US-10-112-944-770	Sequence 770, App
98	6	42.9 <td>597</td> <td>15<th>US-10-276-774-2105</th><th>Sequence 2105, Ap</th></td>	597	15 <th>US-10-276-774-2105</th> <th>Sequence 2105, Ap</th>	US-10-276-774-2105	Sequence 2105, Ap
99	6	42.9 <td>604</td> <td>17<th>US-10-425-115-270800</th><th>Sequence 270800, A</th></td>	604	17 <th>US-10-425-115-270800</th> <th>Sequence 270800, A</th>	US-10-425-115-270800	Sequence 270800, A
100	6	42.9 <td>618</td> <td>15<th>US-10-425-114-62636</th><th>Sequence 62636, A</th></td>	618	15 <th>US-10-425-114-62636</th> <th>Sequence 62636, A</th>	US-10-425-114-62636	Sequence 62636, A
101	6	42.9 <td>625</td> <td>14<th>US-10-029-347-2</th><th>Sequence 2, Appl1</th></td>	625	14 <th>US-10-029-347-2</th> <th>Sequence 2, Appl1</th>	US-10-029-347-2	Sequence 2, Appl1
102	6	42.9 <td>633</td> <td>14<th>US-10-371-905A-24</th><th>Sequence 24, Appl</th></td>	633	14 <th>US-10-371-905A-24</th> <th>Sequence 24, Appl</th>	US-10-371-905A-24	Sequence 24, Appl
103	6	42.9 <td>637</td> <td>15<th>US-10-282-122A-48281</th><th>Sequence 48281, A</th></td>	637	15 <th>US-10-282-122A-48281</th> <th>Sequence 48281, A</th>	US-10-282-122A-48281	Sequence 48281, A
104	6	42.9 <td>642</td> <td>17<th>US-10-425-115-270798</th><th>Sequence 270798, A</th></td>	642	17 <th>US-10-425-115-270798</th> <th>Sequence 270798, A</th>	US-10-425-115-270798	Sequence 270798, A
105	6	42.9 <td>650</td> <td>15<th>US-10-282-122A-66928</th><th>Sequence 66928, A</th></td>	650	15 <th>US-10-282-122A-66928</th> <th>Sequence 66928, A</th>	US-10-282-122A-66928	Sequence 66928, A
106	6	42.9 <td>651</td> <td>9<th>US-09-815-242-11022</th><th>Sequence 11022, A</th></td>	651	9 <th>US-09-815-242-11022</th> <th>Sequence 11022, A</th>	US-09-815-242-11022	Sequence 11022, A
107	6	42.9 <td>651</td> <td>15<th>US-10-282-122A-58169</th><th>Sequence 58169, A</th></td>	651	15 <th>US-10-282-122A-58169</th> <th>Sequence 58169, A</th>	US-10-282-122A-58169	Sequence 58169, A
108	6	42.9 <td>659</td> <td>15<th>US-10-407-866-91</th><th>Sequence 91, Appl</th></td>	659	15 <th>US-10-407-866-91</th> <th>Sequence 91, Appl</th>	US-10-407-866-91	Sequence 91, Appl
109	6	42.9 <td>661</td> <td>15<th>US-10-643-434-7</th><th>Sequence 7, Appl1</th></td>	661	15 <th>US-10-643-434-7</th> <th>Sequence 7, Appl1</th>	US-10-643-434-7	Sequence 7, Appl1
110	6	42.9 <td>676</td> <td>15<th>US-10-092-900A-70</th><th>Sequence 70, Appl</th></td>	676	15 <th>US-10-092-900A-70</th> <th>Sequence 70, Appl</th>	US-10-092-900A-70	Sequence 70, Appl
111	6	42.9 <td>693</td> <td>15<th>US-10-282-122A-78018</th><th>Sequence 78018, A</th></td>	693	15 <th>US-10-282-122A-78018</th> <th>Sequence 78018, A</th>	US-10-282-122A-78018	Sequence 78018, A
112	6	42.9 <td>750</td> <td>16<th>US-10-437-863-129080</th><th>Sequence 129080, A</th></td>	750	16 <th>US-10-437-863-129080</th> <th>Sequence 129080, A</th>	US-10-437-863-129080	Sequence 129080, A
113	6	42.9 <td>757</td> <td>16<th>US-10-437-863-146837</th><th>Sequence 146837, A</th></td>	757	16 <th>US-10-437-863-146837</th> <th>Sequence 146837, A</th>	US-10-437-863-146837	Sequence 146837, A
114	6	42.9 <td>767</td> <td>15<th>US-10-424-599-198727</th><th>Sequence 198727, A</th></td>	767	15 <th>US-10-424-599-198727</th> <th>Sequence 198727, A</th>	US-10-424-599-198727	Sequence 198727, A
115	6	42.9 <td>779</td> <td>16<th>US-10-437-863-133669</th><th>Sequence 133669, A</th></td>	779	16 <th>US-10-437-863-133669</th> <th>Sequence 133669, A</th>	US-10-437-863-133669	Sequence 133669, A
116	6	42.9 <td>782</td> <td>15<th>US-10-282-122A-43733</th><th>Sequence 43733, A</th></td>	782	15 <th>US-10-282-122A-43733</th> <th>Sequence 43733, A</th>	US-10-282-122A-43733	Sequence 43733, A
117	6	42.9 <td>804</td> <td>16<th>US-10-437-863-203648</th><th>Sequence 203648, A</th></td>	804	16 <th>US-10-437-863-203648</th> <th>Sequence 203648, A</th>	US-10-437-863-203648	Sequence 203648, A
118	6	42.9 <td>831</td> <td>17<th>US-10-723-860-2937</th><th>Sequence 2937, Ap</th></td>	831	17 <th>US-10-723-860-2937</th> <th>Sequence 2937, Ap</th>	US-10-723-860-2937	Sequence 2937, Ap
119	6	42.9 <td>862</td> <td>10<th>US-08-825-751A-66</th><th>Sequence 66, Appl</th></td>	862	10 <th>US-08-825-751A-66</th> <th>Sequence 66, Appl</th>	US-08-825-751A-66	Sequence 66, Appl
120	6	42.9 <td>891</td> <td>15<th>US-10-112-944-312</th><th>Sequence 312, App</th></td>	891	15 <th>US-10-112-944-312</th> <th>Sequence 312, App</th>	US-10-112-944-312	Sequence 312, App
121	6	42.9 <td>891</td> <td>17<th>US-10-149-826-2</th><th>Sequence 2, Appl1</th></td>	891	17 <th>US-10-149-826-2</th> <th>Sequence 2, Appl1</th>	US-10-149-826-2	Sequence 2, Appl1
122	6	42.9 <td>892</td> <td>14<th>US-10-446-046-2</th><th>Sequence 2, Appl1</th></td>	892	14 <th>US-10-446-046-2</th> <th>Sequence 2, Appl1</th>	US-10-446-046-2	Sequence 2, Appl1
123	6	42.9 <td>902</td> <td>15<th>US-10-407-866-66</th><th>Sequence 66, Appl</th></td>	902	15 <th>US-10-407-866-66</th> <th>Sequence 66, Appl</th>	US-10-407-866-66	Sequence 66, Appl
124	6	42.9 <td>909</td> <td>14<th>US-10-369-493-23150</th><th>Sequence 23150, A</th></td>	909	14 <th>US-10-369-493-23150</th> <th>Sequence 23150, A</th>	US-10-369-493-23150	Sequence 23150, A
125	6	42.9 <td>947</td> <td>10<th>US-09-865-621-18</th><th>Sequence 18, Appl</th></td>	947	10 <th>US-09-865-621-18</th> <th>Sequence 18, Appl</th>	US-09-865-621-18	Sequence 18, Appl
126	6	42.9 <td>947</td> <td>15<th>US-10-407-866-18</th><th>Sequence 18, Appl</th></td>	947	15 <th>US-10-407-866-18</th> <th>Sequence 18, Appl</th>	US-10-407-866-18	Sequence 18, Appl
127	6	42.9 <td>947</td> <td>16<th>US-10-781-294-18</th><th>Sequence 18, Appl</th></td>	947	16 <th>US-10-781-294-18</th> <th>Sequence 18, Appl</th>	US-10-781-294-18	Sequence 18, Appl
128	6	42.9 <td>1000</td> <td>16<th>US-10-437-863-187544</th><th>Sequence 187544, A</th></td>	1000	16 <th>US-10-437-863-187544</th> <th>Sequence 187544, A</th>	US-10-437-863-187544	Sequence 187544, A
129	6	42.9 <td>1016</td> <td>14<th>US-10-371-905A-4</th><th>Sequence 4, Appl1</th></td>	1016	14 <th>US-10-371-905A-4</th> <th>Sequence 4, Appl1</th>	US-10-371-905A-4	Sequence 4, Appl1
130	6	42.9 <td>1021</td> <td>14<th>US-10-371-905A-8</th><th>Sequence 8, Appl1</th></td>	1021	14 <th>US-10-371-905A-8</th> <th>Sequence 8, Appl1</th>	US-10-371-905A-8	Sequence 8, Appl1
131	5	35.7 <td>9</td> <td>10<th>US-09-935-384-211</th><th>Sequence 211, App</th></td>	9	10 <th>US-09-935-384-211</th> <th>Sequence 211, App</th>	US-09-935-384-211	Sequence 211, App
132	5	35.7 <td>9</td> <td>10<th>US-09-935-384-322</th><th>Sequence 322, App</th></td>	9	10 <th>US-09-935-384-322</th> <th>Sequence 322, App</th>	US-09-935-384-322	Sequence 322, App
133	5	35.7 <td>9</td> <td>10<th>US-09-935-384-326</th><th>Sequence 326, App</th></td>	9	10 <th>US-09-935-384-326</th> <th>Sequence 326, App</th>	US-09-935-384-326	Sequence 326, App
134	5	35.7 <td>9</td> <td>10<th>US-09-935-384-416</th><th>Sequence 416, App</th></td>	9	10 <th>US-09-935-384-416</th> <th>Sequence 416, App</th>	US-09-935-384-416	Sequence 416, App
135	5	35.7 <td>9</td> <td>10<th>US-09-935-384-527</th><th>Sequence 527, App</th></td>	9	10 <th>US-09-935-384-527</th> <th>Sequence 527, App</th>	US-09-935-384-527	Sequence 527, App
136	5	35.7 <td>9</td> <td>10<th>US-09-935-384-610</th><th>Sequence 610, App</th></td>	9	10 <th>US-09-935-384-610</th> <th>Sequence 610, App</th>	US-09-935-384-610	Sequence 610, App
137	5	35.7 <td>10</td> <td>10<th>US-08-572-404B-3732</th><th>Sequence 3732, Ap</th></td>	10	10 <th>US-08-572-404B-3732</th> <th>Sequence 3732, Ap</th>	US-08-572-404B-3732	Sequence 3732, Ap
138	5	35.7 <td>10</td> <td>10<th>US-08-572-404B-3733</th><th>Sequence 3733, Ap</th></td>	10	10 <th>US-08-572-404B-3733</th> <th>Sequence 3733, Ap</th>	US-08-572-404B-3733	Sequence 3733, Ap
139	5	35.7 <td>10</td> <td>10<th>US-08-572-404B-3734</th><th>Sequence 3734, Ap</th></td>	10	10 <th>US-08-572-404B-3734</th> <th>Sequence 3734, Ap</th>	US-08-572-404B-3734	Sequence 3734, Ap
140	5	35.7 <td>10</td> <td>10<th>US-09-935-384-161</th><th>Sequence 161, App</th></td>	10	10 <th>US-09-935-384-161</th> <th>Sequence 161, App</th>	US-09-935-384-161	Sequence 161, App
141	5	35.7 <td>10</td> <td>10<th>US-09-935-384-175</th><th>Sequence 175, App</th></td>	10	10 <th>US-09-935-384-175</th> <th>Sequence 175, App</th>	US-09-935-384-175	Sequence 175, App
142	5	35.7 <td>10</td> <td>10<th>US-09-935-384-372</th><th>Sequence 372, App</th></td>	10	10 <th>US-09-935-384-372</th> <th>Sequence 372, App</th>	US-09-935-384-372	Sequence 372, App
143	5	35.7 <td>10</td> <td>10<th>US-09-935-384-396</th><th>Sequence 396, App</th></td>	10	10 <th>US-09-935-384-396</th> <th>Sequence 396, App</th>	US-09-935-384-396	Sequence 396, App
144	5	35.7 <td>10</td> <td>10<th>US-09-935-384-593</th><th>Sequence 593, App</th></td>	10	10 <th>US-09-935-384-593</th> <th>Sequence 593, App</th>	US-09-935-384-593	Sequence 593, App
145	5	35.7 <td>10</td> <td>10<th>US-09-935-384-692</th><th>Sequence 692, App</th></td>	10	10 <th>US-09-935-384-692</th> <th>Sequence 692, App</th>	US-09-935-384-692	Sequence 692, App
146	5	35.7 <td>12</td> <td>9<th>US-09-071-838-124</th><th>Sequence 124, App</th></td>	12	9 <th>US-09-071-838-124</th> <th>Sequence 124, App</th>	US-09-071-838-124	Sequence 124, App
147	5	35.7 <td>12</td> <td>9<th>US-10-213-512-124</th><th>Sequence 124, App</th></td>	12	9 <th>US-10-213-512-124</th> <th>Sequence 124, App</th>	US-10-213-512-124	Sequence 124, App
148	5	35.7 <td>14</td> <td>17<th>US-10-413-600-28</th><th>Sequence 28, Appl</th></td>	14	17 <th>US-10-413-600-28</th> <th>Sequence 28, Appl</th>	US-10-413-600-28	Sequence 28, Appl
149	5	35.7 <td>16</td> <td>9<th>US-09-841-133-251</th><th>Sequence 251, App</th></td>	16	9 <th>US-09-841-133-251</th> <th>Sequence 251, App</th>	US-09-841-133-251	Sequence 251, App
150	5	35.7 <td>16</td> <td>17<th>US-10-872-155-251</th><th>Sequence 251, App</th></td>	16	17 <th>US-10-872-155-251</th> <th>Sequence 251, App</th>	US-10-872-155-251	Sequence 251, App

## ALIGNMENTS

RESULT 1  
US-10-719-385-20  
Sequence 20, Application US/10719385  
Publication No. US20040209284A1

GENERAL INFORMATION:  
APPLICANT: O'Toole et al.  
TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis  
FILE REFERENCE: 22058-582  
CURRENT APPLICATION NUMBER: US/10/719,385  
CURRENT FILING DATE: 2003-11-21  
PRIOR APPLICATION NUMBER: PCT/US03/37339  
PRIOR FILING DATE: 2003-11-21  
PRIOR APPLICATION NUMBER: 60/428,094  
PRIOR FILING DATE: 2002-11-21  
NUMBER OF SEQ ID NOS: 26  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 20  
LENGTH: 14  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-719-385-20

Query Match 100.0%; Score 14; DB 17; Length 14;  
Best Local Similarity 100.0%; Pred. No. 3.5e-07;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MIRSKITSVLSFC 14  
Db 1 MIRSKITSVLSFC 14

RESULT 2  
US-10-719-385-5  
Sequence 5, Application US/10719385  
Publication No. US20040209284A1  
GENERAL INFORMATION:

APPLICANT: O'Toole et al.  
TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis  
FILE REFERENCE: 22058-582  
CURRENT APPLICATION NUMBER: US/10/719,385  
CURRENT FILING DATE: 2003-11-21  
PRIOR APPLICATION NUMBER: PCT/US03/37339  
PRIOR FILING DATE: 2003-11-21  
PRIOR APPLICATION NUMBER: 60/428,094  
PRIOR FILING DATE: 2002-11-21  
NUMBER OF SEQ ID NOS: 26  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 5  
LENGTH: 1752  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-719-385-5

Query Match 100.0%; Score 14; DB 17; Length 1752;  
Best Local Similarity 100.0%; Pred. No. 1.9e-05;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MIRSKITSVLSFC 14  
Db 1 MIRSKITSVLSFC 14

RESULT 3  
US-10-719-385-2  
Sequence 2, Application US/10719385  
Publication No. US20040209284A1  
GENERAL INFORMATION:  
APPLICANT: O'Toole et al.  
TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis  
FILE REFERENCE: 22058-582  
CURRENT APPLICATION NUMBER: US/10/719,385  
CURRENT FILING DATE: 2003-11-21  
PRIOR APPLICATION NUMBER: PCT/US03/37339  
PRIOR FILING DATE: 2003-11-21  
PRIOR APPLICATION NUMBER: 60/428,094  
PRIOR FILING DATE: 2002-11-21  
NUMBER OF SEQ ID NOS: 26

```
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 1753
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-385-2
```

```
Query Match          100.0%; Score 14; DB 17; Length 1753;
Best Local Similarity 100.0%; Pred. No. 1.9e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Oy      1 MIRSKITSVLSFC 14
        |||||
        1 MIRSKITSVLSFC 14
```

```
RESULT 4
US-10-719-385-3
; Sequence 3, Application US/10719385
; Publication No. US20040209284A1
; GENERAL INFORMATION:
```

```
; APPLICANT: O'Toole et al.
; TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis
; FILE REFERENCE: 22058-582
; CURRENT APPLICATION NUMBER: US/10/719,385
; CURRENT FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: PCT/US03/37339
; PRIOR FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: 60/428,094
; PRIOR FILING DATE: 2002-11-21
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 1753
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-385-3
```

```
Query Match          100.0%; Score 14; DB 17; Length 1753;
Best Local Similarity 100.0%; Pred. No. 1.9e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Oy      1 MIRSKITSVLSFC 14
        |||||
        1 MIRSKITSVLSFC 14
```

```
RESULT 5
US-10-719-385-4
; Sequence 4, Application US/10719385
; Publication No. US20040209284A1
; GENERAL INFORMATION:
```

```
; APPLICANT: O'Toole et al.
; TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis
; FILE REFERENCE: 22058-582
; CURRENT APPLICATION NUMBER: US/10/719,385
; CURRENT FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: PCT/US03/37339
; PRIOR FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: 60/428,094
; PRIOR FILING DATE: 2002-11-21
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 1753
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-385-4
```

```
Query Match          100.0%; Score 14; DB 17; Length 1753;
Best Local Similarity 100.0%; Pred. No. 1.9e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Oy      1 MIRSKITSVLSFC 14
        |||||
        1 MIRSKITSVLSFC 14
```

```
RESULT 6
US-10-719-385-6
; Sequence 6, Application US/10719385
; Publication No. US20040209284A1
; GENERAL INFORMATION:
```

```
; APPLICANT: O'Toole et al.
; TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis
; FILE REFERENCE: 22058-582
; CURRENT APPLICATION NUMBER: US/10/719,385
; CURRENT FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: PCT/US03/37339
; PRIOR FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: 60/428,094
; PRIOR FILING DATE: 2002-11-21
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 1753
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-385-6
```

```
Query Match          100.0%; Score 14; DB 17; Length 1753;
Best Local Similarity 100.0%; Pred. No. 1.9e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Oy      1 MIRSKITSVLSFC 14
        |||||
        1 MIRSKITSVLSFC 14
```

```
RESULT 7
US-10-719-385-7
; Sequence 7, Application US/10719385
; Publication No. US20040209284A1
; GENERAL INFORMATION:
```

```
; APPLICANT: O'Toole et al.
; TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis
; FILE REFERENCE: 22058-582
; CURRENT APPLICATION NUMBER: US/10/719,385
; CURRENT FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: PCT/US03/37339
; PRIOR FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: 60/428,094
; PRIOR FILING DATE: 2002-11-21
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 1753
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-385-7
```

```
Query Match          100.0%; Score 14; DB 17; Length 1753;
Best Local Similarity 100.0%; Pred. No. 1.9e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Oy      1 MIRSKITSVLSFC 14
        |||||
        1 MIRSKITSVLSFC 14
```

```
RESULT 8
US-10-719-385-8
; Sequence 8, Application US/10719385
; Publication No. US20040209284A1
; GENERAL INFORMATION:
```

```
; APPLICANT: O'Toole et al.
```

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

Query Match	100.0%	Score 14	DB 17	Length 1753
Best Local Similarity	100.0%	Pred. No.	1.9e-05	
Matches	14	Conservative	0	Mismatches 0
				Indels 0
				Gaps 0
Qy	1	MIRKSKITSVLSFC	14	
Db	1	MIRKSKITSVLSFC	14	

```

RESULT 9
US-10-719-385-9
; Sequence 9, Application US/10719385
; Publication No. US20040209284A1
; GENERAL INFORMATION:
; APPLICANT: O'toole et al.
; TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis
; FILE REFERENCE: 22058-582
; CURRENT APPLICATION NUMBER: US/10/719,385
; CURRENT FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: PCT/US03/37339
; PRIOR FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: 60/428,094
; PRIOR FILING DATE: 2002-11-21
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 9
; LENGTH: 1753
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-385-9

```

Query Match	100.0%;	Score 14;	DB 17;	Length 1753;
Best Local Similarity	100.0%;	Pred. No. 1.9e-05;		
Matches	14;	Conservative	0;	Mismatches 0;
				Indels 0;
				Gaps 0;
Qy	1	MIRSKITSVLSFC	14	
Db	1	MIRSKITSVLSFC	14	

RESULT 10  
 US-10-719-385-10  
 ; Sequence 10, Application US/10719385  
 ; Publication No. US20040209284A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: O'Toole et al.  
 ; TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis  
 ; FILE REFERENCE: 22058-582  
 ; CURRENT APPLICATION NUMBER: 2003-11-21  
 ; CURRENT FILING DATE: 2003-11-21  
 ; PRIOR APPLICATION NUMBER: PCT/US03/37339  
 ; PRIOR FILING DATE: 2003-11-21  
 ; PRIOR APPLICATION NUMBER: 60/428, 094  
 ; PRIOR FILING DATE: 2002-11-21  
 ; NUMBER OF SEQ ID NOS: 26  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 10

```

; LENGTH: 1753
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-385-10

```

	Query Match	Similarity	Score	DB	Length
Best Local Match	100.0%	100.0%	14	17	1753
Pred. No.	1.9e-05				
Matches	14	Conservative	0	Mismatches	0
				Indels	0
				Gaps	0

```

RESULT 11
US-10-719-385-11
: Sequence 11, Application US/10719385
: Publication No. US20040209284A1
: GENERAL INFORMATION:
: APPLICANT: O'Toole et al.
: TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis
: FILE REFERENCE: 22058-582
: CURRENT APPLICATION NUMBER: US/10/719,385
: CURRENT FILING DATE: 2003-11-21
: PRIOR APPLICATION NUMBER: PCT/US03/37339
: PRIOR FILING DATE: 2003-11-21
: PRIOR APPLICATION NUMBER: 60/428,094
: PRIOR FILING DATE: 2002-11-21
: NUMBER OF SEQ ID NOS: 26
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 11
: LENGTH: 1753
: TYPE: PRT
: ORGANISM: Homo sapiens
US-10-719-385-11

```

	Query Match	Similarity	Score	DB	length
Best Local	100.0%	100.0%	1.9e-05	1753	
Matches	14	Conservative	0	Mismatches	0
				Indels	0
				Gaps	0
Qx	1	MIRSKITSVLSFC	14		
Db	1	MIRSKITSVLSFC	14		

```

RESULT 12
US-10-719-385-12
: Sequence 12, Application US/10719385
: Publication No. US20040209284A1
: GENERAL INFORMATION:
: APPLICANT: O'Toole et al.
: TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis
: FILE REFERENCE: 22058-582
: CURRENT APPLICATION NUMBER: US/10/719,385
: CURRENT FILING DATE: 2003-11-21
: PRIOR APPLICATION NUMBER: PCT/US03/37339
: PRIOR FILING DATE: 2003-11-21
: PRIOR APPLICATION NUMBER: 60/428,094
: PRIOR FILING DATE: 2002-11-21
: NUMBER OF SEQ ID NOS: 26
: SOFTWARE: PatentIn Ver. 2.1
: SEQ ID NO 12
: LENGTH: 1753
: TYPE: PRT
: ORGANISM: Homo sapiens
US-10-719-385-12

```

Query Match	100.0%;	Score 14;	DB 17;	length 1753;
Best Local Similarity	100.0%;	Pred. NO. 1.9e-05;		
Matches 14;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
1. MIRSKITSVLSFC 14				

```
Db      1 MIRSKITSVLSFC 14

RESULT 13
US-10-719-385-13
; Sequence 13, Application US/10719385
; Publication No. US20040209284A1
; GENERAL INFORMATION:
; APPLICANT: O'Toole et al.
; TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis
; FILE REFERENCE: 22058-582
; CURRENT APPLICATION NUMBER: US/10/719,385
; CURRENT FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: PCT/US03/37339
; PRIOR FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: 60/428,094
; PRIOR FILING DATE: 2002-11-21
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 1753
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-385-13

Query Match      100.0%; Score 14; DB 17; Length 1753;
Best Local Similarity 100.0%; Pred. No. 1.9e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 MIRSKITSVLSFC 14
        |||||
        1 MIRSKITSVLSFC 14

RESULT 14
US-10-719-385-14
; Sequence 14, Application US/10719385
; Publication No. US20040209284A1
; GENERAL INFORMATION:
; APPLICANT: O'Toole et al.
; TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis
; FILE REFERENCE: 22058-582
; CURRENT APPLICATION NUMBER: US/10/719,385
; CURRENT FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: PCT/US03/37339
; PRIOR FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: 60/428,094
; PRIOR FILING DATE: 2002-11-21
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; LENGTH: 1753
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-385-14

Query Match      100.0%; Score 14; DB 17; Length 1753;
Best Local Similarity 100.0%; Pred. No. 1.9e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 MIRSKITSVLSFC 14
        |||||
        1 MIRSKITSVLSFC 14

RESULT 15
US-10-719-385-15
; Sequence 15, Application US/10719385
; Publication No. US20040209284A1
; GENERAL INFORMATION:
; APPLICANT: O'Toole et al.
; TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis
; FILE REFERENCE: 22058-582

Db      1 MIRSKITSVLSFC 14

CURRENT APPLICATION NUMBER: US/10/719,385
; CURRENT FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: PCT/US03/37339
; PRIOR FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: 60/428,094
; PRIOR FILING DATE: 2002-11-21
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 15
; LENGTH: 1753
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-385-15

Query Match      100.0%; Score 14; DB 17; Length 1753;
Best Local Similarity 100.0%; Pred. No. 1.9e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 MIRSKITSVLSFC 14
        |||||
        1 MIRSKITSVLSFC 14

RESULT 16
US-10-719-385-16
; Sequence 16, Application US/10719385
; Publication No. US20040209284A1
; GENERAL INFORMATION:
; APPLICANT: O'Toole et al.
; TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis
; FILE REFERENCE: 22058-582
; CURRENT APPLICATION NUMBER: US/10/719,385
; CURRENT FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: PCT/US03/37339
; PRIOR FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: 60/428,094
; PRIOR FILING DATE: 2002-11-21
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 16
; LENGTH: 1753
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-385-16

Query Match      100.0%; Score 14; DB 17; Length 1753;
Best Local Similarity 100.0%; Pred. No. 1.9e-05;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 MIRSKITSVLSFC 14
        |||||
        1 MIRSKITSVLSFC 14

RESULT 17
US-10-719-385-17
; Sequence 17, Application US/10719385
; Publication No. US20040209284A1
; GENERAL INFORMATION:
; APPLICANT: O'Toole et al.
; TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis
; FILE REFERENCE: 22058-582
; CURRENT APPLICATION NUMBER: US/10/719,385
; CURRENT FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: PCT/US03/37339
; PRIOR FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: 60/428,094
; PRIOR FILING DATE: 2002-11-21
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 17
; LENGTH: 1753
; TYPE: PRT
```

ORGANISM: Homo sapiens  
US-10-719-385-17

Query Match 100.0%; Score 14; DB 17; Length 1753;  
Best Local Similarity 100.0%; Pred. No. 1.9e-05;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MIRSKITSVLSFC 14  
DB 1 MIRSKITSVLSFC 14

RESULT 18  
US-10-719-385-18  
Sequence 18, Application US/10719385  
Publication No. US20040209284A1  
GENERAL INFORMATION:  
APPLICANT: O'Toole et al.  
TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis  
FILE REFERENCE: 22058-582  
CURRENT FILING DATE: 2003-11-21  
PRIOR APPLICATION NUMBER: PCT/US03/37339  
PRIOR FILING DATE: 2003-11-21  
PRIOR APPLICATION NUMBER: 60/428,094  
PRIOR FILING DATE: 2002-11-21  
NUMBER OF SEQ ID NOS: 26  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 18  
LENGTH: 1753  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-719-385-18

Query Match 100.0%; Score 14; DB 17; Length 1753;  
Best Local Similarity 100.0%; Pred. No. 1.9e-05;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MIRSKITSVLSFC 14  
DB 1 MIRSKITSVLSFC 14

RESULT 19  
US-10-719-385-19  
Sequence 19, Application US/10719385  
Publication No. US20040209284A1  
GENERAL INFORMATION:  
APPLICANT: O'Toole et al.  
TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis  
FILE REFERENCE: 22058-582  
CURRENT FILING DATE: 2003-11-21  
PRIOR APPLICATION NUMBER: PCT/US03/37339  
PRIOR FILING DATE: 2003-11-21  
PRIOR APPLICATION NUMBER: 60/428,094  
PRIOR FILING DATE: 2002-11-21  
NUMBER OF SEQ ID NOS: 26  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 19  
LENGTH: 1753  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-719-385-19

Query Match 100.0%; Score 14; DB 17; Length 1753;  
Best Local Similarity 100.0%; Pred. No. 1.9e-05;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MIRSKITSVLSFC 14  
DB 1 MIRSKITSVLSFC 14

RESULT 20  
US-10-424-599-148727  
Sequence 148727, Application US/10424599  
Publication No. US20040031072A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa Thomas J  
APPLICANT: Kovalic David K  
APPLICANT: Zhou Yihua  
APPLICANT: Cao Yongwei

TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21(53223) B  
CURRENT APPLICATION NUMBER: US/10/424,599  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 285684  
SEQ ID NO 148727  
LENGTH: 165  
TYPE: PRT  
ORGANISM: Glycine max  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_105322C.1.pep  
US-10-424-599-148727

Query Match 50.0%; Score 7; DB 15; Length 165;  
Best Local Similarity 100.0%; Pred. No. 21;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 5 SKITSVL 11  
DB 17 SKITSVL 23

RESULT 21  
US-10-781-014-694  
Sequence 694, Application US/10781014  
Publication No. US20040180408A1  
GENERAL INFORMATION:  
APPLICANT: Pompejus, Markus  
APPLICANT: Kroger, Burkhard  
APPLICANT: Schroder, Hartwig  
APPLICANT: Zeider, Oskar  
APPLICANT: Habermann, Gregor  
TITLE OF INVENTION: CORYNEBACTERIUM GLUTAMICUM GENES ENCODING PROTEINS  
TITLE OF INVENTION: INVOLVED IN CARBON METABOLISM AND ENERGY  
FILE REFERENCE: BGI-126CPCN  
CURRENT APPLICATION NUMBER: US/10/781,014  
CURRENT FILING DATE: 2004-02-17  
PRIOR APPLICATION NUMBER: US 09/602,740  
PRIOR FILING DATE: 2000-06-23  
PRIOR APPLICATION NUMBER: 60/141,031  
PRIOR FILING DATE: 1999-06-25  
PRIOR APPLICATION NUMBER: 60/143,208  
PRIOR FILING DATE: 1999-07-09  
PRIOR APPLICATION NUMBER: 60/151,572  
PRIOR FILING DATE: 1999-08-31  
PRIOR APPLICATION NUMBER: DE 19931412.8  
PRIOR FILING DATE: 1999-07-08  
PRIOR APPLICATION NUMBER: DE 19931413.6  
PRIOR FILING DATE: 1999-07-08  
PRIOR APPLICATION NUMBER: DE 19931419.5  
PRIOR FILING DATE: 1999-07-08  
PRIOR APPLICATION NUMBER: DE 19931420.9  
PRIOR FILING DATE: 1999-07-08  
PRIOR APPLICATION NUMBER: DE 19931424.1  
PRIOR FILING DATE: 1999-07-08  
PRIOR APPLICATION NUMBER: DE 19931428.4  
PRIOR FILING DATE: 1999-07-08  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 784  
SEQ ID NO 694  
LENGTH: 285

TYPE: PRT  
ORGANISM: Corynebacterium glutamicum  
US-10-781-014-694

Query Match 50.0%; Score 7; DB 16; Length 285;  
Best Local Similarity 100.0%; Pred. No. 33;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ITSVLSF 13  
Db 139 ITSVLSF 145

RESULT 22  
US-10-425-114-47775

Sequence 47775, Application US/10425114  
Publication No. US20040034888A1  
GENERAL INFORMATION:  
APPLICANT: Liu, Jingdong  
APPLICANT: Zhou, Yihua  
APPLICANT: Kovalic, David K.  
APPLICANT: Screen, Steven E.  
APPLICANT: Tabaska, Jack E.  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(5313)B  
CURRENT APPLICATION NUMBER: US/10/425,114  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 73128  
SEQ ID NO 47775  
LENGTH: 290  
TYPE: PRT  
ORGANISM: Zea mays  
FEATURE:  
OTHER INFORMATION: Clone ID: LIB3960-022-B11\_Flt.pep  
US-10-425-114-47775

Query Match 50.0%; Score 7; DB 15; Length 290;  
Best Local Similarity 100.0%; Pred. No. 33;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 ITSVLSF 13  
Db 155 ITSVLSF 161

RESULT 23  
US-10-424-599-158692

Sequence 158692, Application US/10424599  
Publication No. US20040031072A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa, Thomas J.  
APPLICANT: Kovalic, David K.  
APPLICANT: Zhou, Yihua  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(5323)B  
CURRENT APPLICATION NUMBER: US/10/424,599  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 285684  
SEQ ID NO 158692  
LENGTH: 363  
TYPE: PRT  
ORGANISM: Glycine max  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_114319C.1.pep  
US-10-424-599-158692

Query Match 50.0%; Score 7; DB 15; Length 363;  
Best Local Similarity 100.0%; Pred. No. 40;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 KITSVLS 12  
Db 135 KITSVLS 141

RESULT 24

US-09-815-242-10735  
Sequence 10735, Application US/09815242  
Patent No. US20020061569A1  
GENERAL INFORMATION:  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Karl L.  
APPLICANT: Zykend, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
FILE REFERENCE: EITRA.011a  
CURRENT APPLICATION NUMBER: US/09/815,242  
CURRENT FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 10735  
LENGTH: 431  
TYPE: PRT  
ORGANISM: Enterococcus faecalis  
US-09-815-242-10735

Query Match 50.0%; Score 7; DB 9; Length 431;  
Best Local Similarity 100.0%; Pred. No. 46;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 6 KITSVLS 12  
Db 398 KITSVLS 404

RESULT 25

US-10-282-122A-57038  
Sequence 57038, Application US/10282122A  
Publication No. US20040029129A1  
GENERAL INFORMATION:  
APPLICANT: Wang, Liangsu  
APPLICANT: Zamudio, Carlos  
APPLICANT: Malone, Cheryl  
APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlsen, Karl  
APPLICANT: Zykend, Judith  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John  
APPLICANT: Carr, Grant  
APPLICANT: Yamamoto, Robert  
APPLICANT: Forsyth, R.  
APPLICANT: Xu, H.  
TITLE OF INVENTION: Identification of Essential Genes in Microorganisms

FILE REFERENCE: ELITRA.034A  
CURRENT APPLICATION NUMBER: US/10/282,122A  
CURRENT FILING DATE: 2003-02-20  
PRIOR APPLICATION NUMBER: 60/191,078  
PRIOR FILING DATE: 2000-03-21  
PRIOR APPLICATION NUMBER: 60/206,848  
PRIOR FILING DATE: 2000-05-23  
PRIOR APPLICATION NUMBER: 60/207,727  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: 60/230,335  
PRIOR FILING DATE: 2000-09-06  
PRIOR APPLICATION NUMBER: 60/230,347  
PRIOR FILING DATE: 2000-09-09  
PRIOR APPLICATION NUMBER: 60/242,578  
PRIOR FILING DATE: 2000-10-23  
PRIOR APPLICATION NUMBER: 60/253,625  
PRIOR FILING DATE: 2000-11-27  
PRIOR APPLICATION NUMBER: 60/257,931  
PRIOR FILING DATE: 2000-12-22  
PRIOR APPLICATION NUMBER: 60/267,636  
PRIOR FILING DATE: 2001-02-09  
PRIOR APPLICATION NUMBER: 60/269,308  
PRIOR FILING DATE: 2001-02-16  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 78614  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 57038  
LENGTH: 431  
TYPE: PRT  
ORGANISM: Enterococcus faecalis  
US-10-282-122A-57038

Query Match 50.0%; Score 7; DB 15; Length 431;  
Best Local Similarity 100.0%; Pred. No. 46;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 KITSVL 12  
|||  
Db 398 KITSVL 404

RESULT 26  
US-09-738-626-6499  
Sequence 6499, Application US/09738626  
Publication No. US20020197605A1  
GENERAL INFORMATION:  
APPLICANT: NAKAGAWA, SATOSHI  
APPLICANT: MIZOGUCHI, HIROSHI  
APPLICANT: ANDO, SEIKO  
APPLICANT: HAYASHI, MIKIRO  
APPLICANT: OCHIAI, KEIKO  
APPLICANT: YOKOI, HARUHIKO  
APPLICANT: YATEISHI, NAOKO  
APPLICANT: SENOH, AKIHIRO  
APPLICANT: IKEDA, MASATO  
APPLICANT: OZAKI, AKIO  
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
FILE REFERENCE: 249-125  
CURRENT APPLICATION NUMBER: US/09/738,626  
PRIOR FILING DATE: 2000-12-18  
PRIOR APPLICATION NUMBER: JP 99/377484  
PRIOR FILING DATE: 1999-12-16  
PRIOR APPLICATION NUMBER: JP 00/159162  
PRIOR FILING DATE: 2000-04-07  
PRIOR APPLICATION NUMBER: JP 00/280988  
PRIOR FILING DATE: 2000-08-03  
NUMBER OF SEQ ID NOS: 7059  
SOFTWARE: PatentIn ver. 3.0  
SEQ ID NO 6499  
LENGTH: 1019  
TYPE: PRT  
ORGANISM: Corynebacterium glutamicum  
US-09-738-626-6499

Query Match 50.0%; Score 7; DB 9; Length 1019;  
Best Local Similarity 100.0%; Pred. No. 93;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 ITSLSF 13  
|||  
Db 139 ITSLSF 145

RESULT 27  
US-10-380-055-12  
Sequence 12, Application US/10380055  
Publication No. US20040014180A1  
GENERAL INFORMATION:  
APPLICANT: BOTT, Michael, et al.  
TITLE OF INVENTION: A Method For Microbial Production Of Metabolic Products, Polynucle  
FILE REFERENCE: 032301 WN 331  
CURRENT APPLICATION NUMBER: US/10/380,055  
CURRENT FILING DATE: 2003-03-11  
NUMBER OF SEQ ID NOS: 35  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 12  
LENGTH: 1019  
TYPE: PRT  
ORGANISM: Corynebacterium glutamicum  
US-10-380-055-12

Query Match 50.0%; Score 7; DB 15; Length 1019;  
Best Local Similarity 100.0%; Pred. No. 93;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 7 ITSLSF 13  
|||  
Db 139 ITSLSF 145

RESULT 28  
US-10-424-599-148728  
Sequence 148728, Application US/10424599  
Publication No. US20040031072A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa, Thomas J  
APPLICANT: Kovalic, David K  
APPLICANT: Zhou, Yihua  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(53223)B  
CURRENT APPLICATION NUMBER: US/10/424,599  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 285684  
SEQ ID NO 148728  
LENGTH: 1065  
TYPE: PRT  
ORGANISM: Glycine max  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_105323C.1.pep  
US-10-424-599-148728

Query Match 50.0%; Score 7; DB 15; Length 1065;  
Best Local Similarity 100.0%; Pred. No. 97;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 5 SKITSVL 11  
|||  
Db 739 SKITSVL 745

RESULT 29  
US-10-075-869-10  
Sequence 10, Application US/10075869



Publication No. US20030104622A1  
GENERAL INFORMATION:  
APPLICANT: Robbins, Paul D.  
APPLICANT: Ml, Zhibao  
APPLICANT: Frizzeil, Raymond  
APPLICANT: Glorioso, Joseph C.  
APPLICANT: Gambotto, Andrea  
TITLE OF INVENTION: IDENTIFICATION OF PEPTIDES THAT  
FACILITATE UPTAKE AND CYTOPLASMIC AND/OR NUCLEAR TRANSPORT  
TITLE OF INVENTION: OF PROTEINS, DNA AND VIRUSES  
FILE REFERENCE: AP32573-AA 072396.0237  
CURRENT FILING DATE: 2002-02-13  
PRIORITY APPLICATION NUMBER: US/10/075,869  
PRIORITY FILING DATE: 1999-09-01  
PRIORITY APPLICATION NUMBER: 60/188,944  
PRIORITY FILING DATE: 2000-03-13  
NUMBER OF SEQ ID NOS: 99  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 10  
LENGTH: 12  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: random peptide library  
US-10-075-869-10

Query Match 42.9%; Score 6; DB 14; Length 12;  
Best Local Similarity 100.0%; Pred. No. 23;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 IRKSKI 7  
Db 3 IRKSKI 8

## RESULT 30

US-10-366-493-10  
Sequence 10, Application US/10366493  
Publication No. US20030219826A1  
GENERAL INFORMATION:  
APPLICANT: Robbins, Paul D.  
APPLICANT: Ml, Zhibao  
APPLICANT: Frizzeil, Raymond  
APPLICANT: Glorioso, Joseph C.  
APPLICANT: Gambotto, Andrea  
TITLE OF INVENTION: IDENTIFICATION OF PEPTIDES THAT FACILITATE UPTAKE AND CYTOPLASMIC  
TRANSPORT  
TITLE OF INVENTION: OF PROTEINS, DNA AND VIRUSES  
FILE REFERENCE: AP32573-A-A-A 072396.0246  
CURRENT APPLICATION NUMBER: US/10/366,493  
CURRENT FILING DATE: 2003-02-12  
PRIORITY APPLICATION NUMBER: 10/075,869  
PRIORITY FILING DATE: 2002-02-13  
PRIORITY APPLICATION NUMBER: 09/653,182  
PRIORITY FILING DATE: 2000-08-31  
PRIORITY APPLICATION NUMBER: 60/188,944  
PRIORITY FILING DATE: 2000-03-13  
PRIORITY APPLICATION NUMBER: 60/151,980  
PRIORITY FILING DATE: 1999-09-01  
NUMBER OF SEQ ID NOS: 107  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 10  
LENGTH: 12  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: pep 10  
US-10-366-493-10

Query Match 42.9%; Score 6; DB 14; Length 12;  
Best Local Similarity 100.0%; Pred. No. 23;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 2 IRKSKI 7  
Db 3 IRKSKI 8

## RESULT 31

US-10-029-347-10  
Sequence 10, Application US/10029347  
Publication No. US20030017562A1  
GENERAL INFORMATION:  
APPLICANT: Bristol-Myers Squibb Company  
TITLE OF INVENTION: A NOVEL HUMAN LEUCINE-RICH REPEAT CONTAINING PROTEIN EXPRESSED  
IN SMALL INTESTINE, HMR511  
FILE REFERENCE: D0066  
CURRENT APPLICATION NUMBER: US/10/029,347  
CURRENT FILING DATE: 2001-12-20  
NUMBER OF SEQ ID NOS: 28  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 10  
LENGTH: 17  
TYPE: PRT  
ORGANISM: homo sapiens  
US-10-029-347-10

Query Match 42.9%; Score 6; DB 14; Length 17;  
Best Local Similarity 100.0%; Pred. No. 31;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 ITSVLS 12  
Db 7 ITSVLS 12

## RESULT 32

US-10-437-963-111764  
Sequence 111764, Application US/10437963  
Publication No. US20040123343A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa, Thomas J.  
APPLICANT: Kovalic, David K.  
APPLICANT: Zhou, Yihua  
APPLICANT: Cao, Yongwei  
APPLICANT: Wu, Wei  
APPLICANT: Bouharov, Andrey A.  
APPLICANT: Barbazuk, Brad  
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21(53221)B  
CURRENT APPLICATION NUMBER: US/10/437,963  
CURRENT FILING DATE: 2003-05-14  
NUMBER OF SEQ ID NOS: 204966  
SEQ ID NO 111764  
LENGTH: 48  
TYPE: PRT  
ORGANISM: Oryza sativa  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT4530\_15711C.1.pep  
US-10-437-963-111764

Query Match 42.9%; Score 6; DB 16; Length 48;  
Best Local Similarity 100.0%; Pred. No. 72;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 9 SVLSFC 14  
Db 40 SVLSFC 45

## RESULT 33

US-10-424-599-267615

```
; Sequence 267615, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 267615
; LENGTH: 53
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_8367C.1.pep
US-10-424-599-267615
```

```
Query Match 42.9%; Score 6; DB 15; Length 53;
Best Local Similarity 100.0%; Pred. No. 78;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 7 ITSVLS 12
DB 3 ITSVLS 8
```

```
RESULT 34
US-10-425-115-185349
; Sequence 185349, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 185349
; LENGTH: 56
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_100626C.1.pep
US-10-425-115-185349
```

```
Query Match 42.9%; Score 6; DB 17; Length 56;
Best Local Similarity 100.0%; Pred. No. 82;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 4 KSKITS 9
DB 47 KSKITS 52
```

```
RESULT 35
US-10-437-963-197926
; Sequence 197926, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
```

```
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 197926
; LENGTH: 60
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_93637C.1.pep
US-10-437-963-197926
```

```
Query Match 42.9%; Score 6; DB 16; Length 60;
Best Local Similarity 100.0%; Pred. No. 87;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 4 KSKITS 9
DB 53 KSKITS 58
```

```
RESULT 36
US-10-424-599-284793
; Sequence 284793, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 284793
; LENGTH: 69
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_99196C.1.pep
US-10-424-599-284793
```

```
Query Match 42.9%; Score 6; DB 15; Length 69;
Best Local Similarity 100.0%; Pred. No. 97;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 5 SKITSV 10
DB 8 SKITSV 13
```

```
RESULT 37
US-10-424-599-143794
; Sequence 143794, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 143794
```

LENGTH: 82  
TYPE: PRT  
ORGANISM: Glycine max  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_10085C.1.pcp  
US-10-424-599-143794

Query Match 42.9%; Score 6; DB 15; Length 82;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 IRKSKI 7  
Db 19 IRKSKI 24

RESULT 38  
US-10-424-599-217068  
Sequence 217068, Application US/10424599  
Publication No. US20040031072A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa, Thomas J  
APPLICANT: Kovalic, David K  
APPLICANT: Zhou, Yihua  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(53223)B  
CURRENT APPLICATION NUMBER: US/10/424,599  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 285684  
SEQ ID NO 217068  
LENGTH: 100  
TYPE: PRT  
ORGANISM: Glycine max  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_3803C.1.pcp  
US-10-424-599-217068

Query Match 42.9%; Score 6; DB 15; Length 100;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 8 TSUUSF 13  
Db 73 TSUUSF 78

RESULT 39  
US-10-425-115-307173  
Sequence 307173, Application US/10425115  
Publication No. US20040214272A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa, Thomas J.  
APPLICANT: Kovalic, David K.  
APPLICANT: Zhou, Yihua  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(53222)B  
CURRENT APPLICATION NUMBER: US/10/425,115  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 369326  
SEQ ID NO 307173  
LENGTH: 100  
TYPE: PRT  
ORGANISM: Zea mays  
FEATURE:  
OTHER INFORMATION: Clone ID: MRT4577\_43212C.1.pcp  
US-10-425-115-307173

Query Match 42.9%; Score 6; DB 17; Length 100;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 5 SKITSV 10  
Db 48 SKITSV 53

RESULT 40  
US-10-425-114-54103  
Sequence 54103, Application US/10425114  
Publication No. US20040034888A1  
GENERAL INFORMATION:  
APPLICANT: Liu, Jingdong  
APPLICANT: Zhou, Yihua  
APPLICANT: Kovalic, David K.  
APPLICANT: Screen, Steven E  
APPLICANT: Tabaska, Jack E  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(53313)B  
CURRENT APPLICATION NUMBER: US/10/425,114  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 73128  
SEQ ID NO 54103  
LENGTH: 111  
TYPE: PRT  
ORGANISM: Zea mays  
FEATURE:  
OTHER INFORMATION: Clone ID: 700577676\_F11.pcp  
US-10-425-114-54103

Query Match 42.9%; Score 6; DB 15; Length 111;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 KITSVL 11  
Db 15 KITSVL 20

Search completed: December 30, 2004, 15:48:49  
Job time : 147 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: December 30, 2004, 14:45:25 ; Search time 0.934622 Seconds  
(without alignment)  
3693.737 Million cell updates/sec

Title: US-10-719-385-26

Perfect score: 36

Sequence: 1 AGGRCV 6

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Database : Uniprot\_02.\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	36	100.0	339	2 Q7XSQ0	Q7XSQ0 oryza sativ
2	36	100.0	407	2 Q9UPF4	Q9UPF4 homo sapien
3	36	100.0	444	1 Y966_METTA	Y966_METTA
4	36	100.0	469	2 O28589	O28589 archaeglob
5	36	100.0	499	2 Q7SYV0	Q7SYV0 xenopus lae
6	36	100.0	798	2 Q8C7A1	Q8C7A1 mus musculu
7	36	100.0	843	2 Q6GQ09	Q6GQ09 homo sapien
8	36	100.0	858	2 Q9N0S3	Q9N0S3 homo sapien
9	36	100.0	1639	2 Q7Z3K8	Q7Z3K8 homo sapien
10	36	100.0	1745	2 Q14675	Q14675 homo sapien
11	36	100.0	1761	2 Q6Z0H8	Q6Z0H8 mus musculu
12	36	100.0	1761	2 BAC97883	BAC97883 mus muscu
13	35	97.2	324	2 Q8S345	Q8S345 capsicum an
14	35	97.2	383	2 Q50225	Q50225 methanosarc
15	35	97.2	383	2 Q50248	Q50248 methanosarc
16	35	97.2	386	2 Q8PUW0	Q8PUW0 methanosarc
17	35	97.2	386	2 Q8PV06	Q8PV06 methanosarc
18	35	97.2	393	2 Q8A8W8	Q8A8W8 bacteroides
19	35	97.2	398	2 Q9W505	Q9W505 vicia vilif
20	35	97.2	409	2 Q93XJ1	Q93XJ1 salix gligi
21	35	97.2	418	2 Q6U7H9	Q6U7H9 malus domes
22	35	97.2	418	2 AAQ84042	AAQ84042 malus dom
23	35	97.2	501	2 Q27917	Q27917 methanobac
24	35	97.2	1036	1 ACR1_HUMAN	ACR1_HUMAN
25	33	91.7	93	2 Q84SX1	Q84SX1 oryza sativ
26	33	91.7	94	2 Q8LH84	Q8LH84 oryza sativ
27	33	91.7	591	1 U149_EBV	U149_EBV
28	33	91.7	591	2 Q777G6	Q777G6 human herpe
29	33	91.7	591	2 CAD53400	CAD53400 human her
30	33	91.7	2434	2 Q6RKJ9	Q6RKJ9 botrytis ci
31	33	91.7	2434	2 AAR90237	AAR90237 botrytis

32	32	88.9	73	2 Q66009	Q66009 bacillus su
33	32	88.9	77	2 Q75WH6	Q75WH6 macrochele
34	32	88.9	77	2 BAD13402	BAD13402 macrochele
35	32	88.9	91	2 Q7UYG4	Q7UYG4 rhodopirelli
36	32	88.9	106	2 Q9CTQ0	Q9CTQ0 mus musculu
37	32	88.9	108	2 Q8BPV6	Q8BPV6 mus musculu
38	32	88.9	119	2 Q7KUM2	Q7KUM2 drosophila
39	32	88.9	119	2 AAS65002	AAS65002 drosophila
40	32	88.9	132	2 Q6SHL1	Q6SHL1 uncultured
41	32	88.9	132	2 AAR37610	AAR37610 uncultured
42	32	88.9	136	2 Q25300	Q25300 leishmania
43	32	88.9	137	2 Q6G3N4	Q6G3N4 bartonella
44	32	88.9	140	2 Q87XK2	Q87XK2 pseudomonas
45	32	88.9	147	2 Q7WY76	Q7WY76 bordetella
46	32	88.9	148	2 Q8LR19	Q8LR19 oryza sativ
47	32	88.9	156	2 Q7ZB17	Q7ZB17 desulfovibr
48	32	88.9	156	2 AAS95242	AAS95242 desulfovi
49	32	88.9	158	2 Q943U7	Q943U7 oryza sativ
50	32	88.9	200	2 Q95HV9	Q95HV9 peromyscus
51	32	88.9	200	2 Q95HW1	Q95HW1 peromyscus
52	32	88.9	200	2 Q9RUT2	Q9RUT2 streptomyce
53	32	88.9	206	2 Q6G3N6	Q6G3N6 bartonella
54	32	88.9	208	2 Q94243	Q94243 schizosacch
55	32	88.9	214	2 Q9F6W0	Q9F6W0 rhizobium e
56	32	88.9	216	2 Q8N7M1	Q8N7M1 homo sapien
57	32	88.9	218	2 Q96NZ0	Q96NZ0 homo sapien
58	32	88.9	229	2 Q8ZEH8	Q8ZEH8 streptomyce
59	32	88.9	247	2 Q7T9U2	Q7T9U2 adoxophyes
60	32	88.9	253	2 Q8N8T0	Q8N8T0 homo sapien
61	32	88.9	254	2 Q81604	Q81604 mesembryant
62	32	88.9	256	2 P72460	P72460 streptomyce
63	32	88.9	266	2 Q8NAJ1	Q8NAJ1 homo sapien
64	32	88.9	269	2 Q7X6S7	Q7X6S7 oryza sativ
65	32	88.9	269	2 CAD41855	CAD41855 oryza sat
66	32	88.9	269	2 CAE05968	CAE05968 oryza sat
67	32	88.9	275	2 Q9ALM7	Q9ALM7 saccharopol
68	32	88.9	280	2 P96821	P96821 mycobacteri
69	32	88.9	280	2 Q7U2R4	Q7U2R4 mycobacteri
70	32	88.9	287	2 Q7XK68	Q7XK68 oryza sativ
71	32	88.9	292	1 PER2_CUCSA	PER2_CUCSA
72	32	88.9	312	2 Q8L442	Q8L442 oryza sativ
73	32	88.9	314	2 Q9NE71	Q9NE71 leishmania
74	32	88.9	320	2 Q982F1	Q982F1 rhizobium l
75	32	88.9	322	2 Q42964	Q42964 nicotiana t
76	32	88.9	323	2 Q6K770	Q6K770 oryza sativ
77	32	88.9	323	2 BAD21872	BAD21872 oryza sat
78	32	88.9	323	2 BAD21877	BAD21877 oryza sat
79	32	88.9	324	2 Q7PKX8	Q7PKX8 anopheles g
80	32	88.9	326	2 Q90XK3	Q90XK3 gnathonevus
81	32	88.9	341	2 Q98VG3	Q98VG3 human herpe
82	32	88.9	342	2 Q6Z135	Q6Z135 oryza sativ
83	32	88.9	342	2 BAD15483	BAD15483 oryza sat
84	32	88.9	352	2 Q82MX5	Q82MX5 streptomyce
85	32	88.9	374	1 SBP_CRYTA	SBP_CRYTA
86	32	88.9	374	2 Q8RUR1	Q8RUR1 cryptomeria
87	32	88.9	382	2 Q8TUD9	Q8TUD9 methanosarc
88	32	88.9	398	2 Q43783	Q43783 musa acumin
89	32	88.9	398	2 Q7V908	Q7V908 prochloroco
90	32	88.9	398	2 Q93OR4	Q93OR4 rhizobium m
91	32	88.9	404	1 PE18_LYCES	PE18_LYCES
92	32	88.9	406	2 Q8T3W6	Q8T3W6 drosophila
93	32	88.9	407	1 EVX1_HUMAN	EVX1_HUMAN
94	32	88.9	407	2 Q95DW4	Q95DW4 musa acumin
95	32	88.9	408	1 P118_ARATH	P118_ARATH
96	32	88.9	423	2 Q93256	Q93256 gallus galli
97	32	88.9	425	1 Y450_HUMAN	Y450_HUMAN
98	32	88.9	426	2 Q97XZ7	Q97XZ7 clostridium
99	32	88.9	432	1 PL22_ARATH	PL22_ARATH
100	32	88.9	437	1 E2F1_HUMAN	E2F1_HUMAN
101	32	88.9	438	2 Q6L168	Q6L168 methanococ
102	32	88.9	438	2 CAF30679	CAF30679 methanoc
103	32	88.9	448	1 FXN4_HUMAN	FXN4_HUMAN
104	32	88.9	450	1 VD10_BPT5	VD10_BPT5

Q66009	bacillus su
Q75WH6	macrochele
BAD13402	macrochele
Q7UYG4	rhodopirelli
Q9CTQ0	mus musculu
Q8BPV6	mus musculu
Q7KUM2	drosophila
AAS65002	drosophila
Q6SHL1	uncultured
AAR37610	uncultured
Q25300	leishmania
Q6G3N4	bartonella
Q87XK2	pseudomonas
Q7WY76	bordetella
Q8LR19	oryza sativ
Q7ZB17	desulfovibr
AAS95242	desulfovi
Q943U7	oryza sativ
Q95HV9	peromyscus
Q95HW1	peromyscus
Q9RUT2	streptomyce
Q6G3N6	bartonella
Q94243	schizosacch
Q9F6W0	rhizobium e
Q8N7M1	homo sapien
Q96NZ0	homo sapien
Q8ZEH8	streptomyce
Q7T9U2	adoxophyes
Q8N8T0	homo sapien
Q81604	mesembryant
P72460	streptomyce
Q8NAJ1	homo sapien
Q7X6S7	oryza sativ
CAD41855	oryza sat
CAE05968	oryza sat
Q9ALM7	saccharopol
P96821	mycobacteri
Q7U2R4	mycobacteri
Q7XK68	oryza sativ
P19135	cucumis bat
Q8L442	oryza sativ
Q9NE71	leishmania
Q982F1	rhizobium l
Q42964	nicotiana t
Q6K770	oryza sativ
BAD21872	oryza sat
BAD21877	oryza sat
Q7PKX8	anopheles g
Q90XK3	gnathonevus
Q98VG3	human herpe
Q6Z135	oryza sativ
BAD15483	oryza sat
Q82MX5	streptomyce
P18632	cryptomeria
Q8RUR1	cryptomeria
Q8TUD9	methanosarc
Q43783	musa acumin
Q7V908	prochloroco
Q93OR4	rhizobium m
PE18_LYCES	lycopersico
Q8T3W6	drosophila
P96400	homo sapien
Q95DW4	musa acumin
Q95DM8	arabidopsis
Q93256	gallus galli
Y450_HUMAN	homo sapien
Q97XZ7	clostridium
Q93225	arabidopsis
Q01094	homo sapien
Q6L168	methanococ
CAF30679	methanococ
Q96NZ0	homo sapien
P1107	bacterioph

105 32 88.9 450 2 AAS7170 bacteriop  
106 32 88.9 453 2 O992A3 O992A3 clostridium  
107 32 88.9 454 2 O969U8 O969U8 homo sapien  
108 32 88.9 454 2 O97GT6 O97GT6 clostridium  
109 32 88.9 455 2 O96CH2 O96CH2 shewanella  
110 32 88.9 455 2 O96T11 O96T11 mus musculu  
111 32 88.9 455 2 O920R1 O920R1 mus musculu  
112 32 88.9 457 2 O96XJ8 O96XJ8 clostridium  
113 32 88.9 462 1 SVG\_TREPA SVG\_TREPA treponema p  
114 32 88.9 462 2 O96H26 O96H26 homo sapien  
115 32 88.9 463 2 O96H60 O96H60 m mus muscu  
116 32 88.9 478 2 O70PA5 O70PA5 melittangiu  
117 32 88.9 478 2 O6N9V7 O6N9V7 rhodopsedu  
118 32 88.9 478 2 CAD89766 CAD89766 melittang  
119 32 88.9 478 2 CAE26872 CAE26872 rhodopsen  
120 32 88.9 479 1 FXLG\_HUMAN FXLG\_HUMAN sapien  
121 32 88.9 497 1 P90338 P90338 saguaro cac  
122 32 88.9 503 2 O61E15 O61E15 oryza sativ  
123 32 88.9 506 2 O628E9 O628E9 oryza sativ  
124 32 88.9 506 2 BAD07870 BAD07870 oryza sat  
125 32 88.9 508 2 O8K234 O8K234 mus musculu  
126 32 88.9 508 2 O9D682 O9D682 mus musculu  
127 32 88.9 513 2 O7PIU8 O7PIU8 chromobacte  
128 32 88.9 531 1 GHT2\_SCHPO GHT2\_SCHPO schizosacch  
129 32 88.9 534 2 O6C9N1 O6C9N1 yarrowia li  
130 32 88.9 535 1 GHT6\_SCHPO GHT6\_SCHPO schizosacch  
131 32 88.9 587 2 O88831 O88831 rattus norv  
132 32 88.9 588 2 O8C078 O8C078 mus musculu  
133 32 88.9 588 2 O8CH42 O8CH42 mus musculu  
134 32 88.9 616 2 O8RBB3 O8RBB3 thermocanaer  
135 32 88.9 617 2 O8X116 O8X116 clostridium  
136 32 88.9 658 2 O94917 O94917 rhodophila  
137 32 88.9 672 2 O7US07 O7US07 rhodospirilli  
138 32 88.9 687 2 O9XWD1 O9XWD1 caenorhabdi  
139 32 88.9 699 1 VGLG\_HHV2H VGLG\_HHV2H human herpe  
140 32 88.9 717 2 O922U5 O922U5 mus musculu  
141 32 88.9 719 2 O82MW5 O82MW5 streptomyce  
142 32 88.9 744 2 O12715 O12715 trichoderma  
143 32 88.9 747 2 O6UJY0 O6UJY0 trichoderma  
144 32 88.9 747 2 AAO76093 AAO76093 trichoder  
145 32 88.9 750 2 O6Z1X0 O6Z1X0 oryza sativ  
146 32 88.9 750 2 BAC92520 BAC92520 oryza sat  
147 32 88.9 755 2 P89107 P89107 saguaro cac  
148 32 88.9 773 2 O6UTU1 O6UTU1 oryza sativ  
149 32 88.9 773 2 AAO56304 AAO56304 oryza sat  
150 32 88.9 791 2 O96L64 O96L64 homo sapien

## ALIGNMENTS

RESULT 1  
Q7XS00 PRELIMINARY; PRT; 339 AA.

AC Q7XS00; 01-OCT-2003 (TRENBLrel. 25, Created)  
DT 01-MAR-2004 (TRENBLrel. 26, Last sequence update)  
DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)  
DE OSJNB0084K11.19 protein.  
GN Name=OSJNB0084K11.19;  
OS Oryza sativa (japonica cultivar-group).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
OC Ehrhartoideae; Oryzaceae; Oryza.  
OX NCBI\_TaxID=33947;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX PubMed=12447439;  
RA Feng Q., Zhang Y., Hao P., Wang S., Fu G., Huang Y., Li Y., Zhu J.,  
RA Liu Y., Hu X., Jia P., Zhang Y., Zhao Q., Yang K., Yu S., Tang Y.,  
RA Weng Q., Zhang L., Lu Y., Mu J., Lu Y., Zhang L.S., Yu Z., Fan D.,  
RA Liu X., Lu T., Li C., Wu Y., Sun T., Lei H., Li T., Hu H., Guan J.,  
RA Wu M., Zhang R., Zhou B., Chen Z., Chen L., Jin Z., Wang R., Yin H.,

RA Cai Z., Ren S., Lv G., Gu W., Zhu G., Tu Y., Jia J., Zhang Y.,  
RA Chen J., Kang H., Chen X., Shao C., Sun Y., Hu O., Zhang X., Zhang W.,  
RA Wang L., Ding C., Sheng H., Gu J., Chen S., Ni L., Zhu F., Chen W.,  
RA Lan L., Lai Y., Cheng Z., Gu W., Jiang J., Li J., Hong G., Xue Y.,  
RA Han B.;  
RT "Sequence and analysis of rice chromosome 4.";  
RL Nature 420:316-320(2002).  
DR EMBL; AL606687; CAE01851.2; -  
DR Gramene; O7XS00; -  
DR GO; GO:0046872; F:metal ion binding; IEA.  
DR InterPro; IPR001594; Znf\_DHHC.  
DR Pfam; PFO1529; Zf-DHHC; I.  
DR ProDom; PD003041; Znf\_DHHC; I.  
DR PROSITE; PS50216; ZF\_DHHC; I.  
SQ SEQUENCE 339 AA; 38946 MW; 56F6B67F2A2650E CRC64;

Query Match 100.0%; Score 36; DB 2; Length 339;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGPCV 6  
DB 6 AGGPCV 11

RESULT 2  
Q9UFP4 PRELIMINARY; PRT; 407 AA.

AC Q9UFP4; 01-MAY-2000 (TRENBLrel. 13, Created)  
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)  
DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)  
DE Hypothetical protein DKFZp434H0717 (Fragment).  
GN Name=DKFZp434H0717;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Testis;  
RA Bioecker H., Boecker M., Brandt P., Mewes H.W., Gassenhuber J.,  
RA Wiemann S.;  
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL122102; CAB59268.1; -  
DR PIR; T34535; T34535.  
KW Hypothetical protein.  
FT NON\_TER 1  
SQ SEQUENCE 407 AA; 43799 MW; FD4F43CE2AC0DB3C CRC64;

Query Match 100.0%; Score 36; DB 2; Length 407;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGPCV 6  
DB 280 AGGPCV 285

## RESULT 3

Y966\_METUA  
ID Y966\_METUA STANDARD; PRT; 444 AA.  
AC O58376;  
DT 29-MAR-2004 (Rel. 43, Created)  
DT 29-MAR-2004 (Rel. 43, Last sequence update)  
DT 05-JUL-2004 (Rel. 44, Last annotation update)  
DE Hypothetical protein MJ0966;  
GN OrderedLocustNames=MJ0966;  
OS Methanococcus jamaeschi.  
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;  
OC Methanocaldococcaceae; Methanocaldococcus.  
OX NCBI\_TaxID=21907;  
RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;  
 RX MEDLINE=96337999; PubMed=8688087;  
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,  
 RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,  
 RA Kertevage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reisch C.I.,  
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodok A.,  
 RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fultman J.L., Nguyen D.,  
 RA Uterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,  
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,  
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;  
 RT "Complete genome sequence of the methanogenic archaeon, Methanococcus  
 jannaschii.";  
 RL Science 273:1058-1073 (1996).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; U67539; AAB89868.1; -.  
 DR PIR; F64420; F64420.  
 DR TIGR; MJ0966; -.  
 DR InterPro; IPR006638; E1p3/M1AB/N1FB.  
 DR InterPro; IPR007197; Radical SAM.  
 DR Pfam; PF04055; Radical SAM; I.  
 DR SMART; SM00729; E1p3; I.  
 KW Complete proteome; Hypothetical protein.  
 SQ SEQUENCE 444 AA; 51347 MW; 25ABAB5890A20C70 CRC64;  
 Query Match 100.0%; Score 36; DB 1; Length 444;  
 Best Local Similarity 100.0%; Pred. No. 1.6e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 1 AGPCV 6  
 Db 88 AGPCV 93  
 RESULT 4  
 ID 028589 PRELIMINARY; PRT; 469 AA.  
 AC 028589;  
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Hypothetical protein AFI684.  
 GN OrderedLocustNames=AFI684;  
 OS Archaeoglobus fulgidus.  
 OC Archaeae; Euryarchaeota; Archaeoglobi; Archaeoglobales;  
 OC Archaeoglobaceae; Archaeoglobus.  
 OX NCBI\_TaxID=2234;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;  
 RX MEDLINE=98049343; PubMed=9389475; DOI=10.1038/37052;  
 RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,  
 RA Ketchum K.A., Dodson R.J., Gwinn M.L., Hickey E.K., Peterson J.D.,  
 RA Richardson D.L., Kertevage A.R., Graham D.E., Kyriakides N.C.,  
 RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.R.,  
 RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.J.,  
 RA Peterson S.N., Reich C.I., McNeil L.K., Badger J.H., Glodok A.,  
 RA Zhou L., Overbeek R., Gocayne J.D., Weidman J.F., McDonald L.A.,  
 RA Uterback T.R., Cotton M.D., Spriggs T., Artiach P., Kaine B.P.,  
 RA Sykes S.M., Sadow P.W., D'Andrea K.P., Bowman C., Fujii C.,  
 RA Garland S.A., Mason T.M., Olsen G.J., Fraser C.M., Smith H.O.,  
 RA Woese C.R., Venter J.C.;  
 RT "The complete genome sequence of the hyperthermophilic, sulphate-  
 RT reducing archaeon Archaeoglobus fulgidus.";  
 RL Nature 390:364-370 (1997).

DR EMBL; AE000987; AAB89565.1; -.  
 DR PIR; C69460; C69460.  
 DR TIGR; AFI684; -.  
 DR InterPro; IPR006638; E1p3/M1AB/N1FB.  
 DR InterPro; IPR007197; Radical SAM.  
 DR Pfam; PF04055; Radical SAM; I.  
 DR SMART; SM00729; E1p3; I.  
 KW Complete proteome; Hypothetical protein.  
 SQ SEQUENCE 469 AA; 53468 MW; 0FE86155B1F1004 CRC64;  
 Query Match 100.0%; Score 36; DB 2; Length 469;  
 Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 1 AGPCV 6  
 Db 104 AGPCV 109  
 RESULT 5  
 ID 075V0 PRELIMINARY; PRT; 499 AA.  
 AC 075V0;  
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE Parsl-prov protein.  
 OS Xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;  
 OC Xenopodinae; Xenopus.  
 OX NCBI\_TaxID=8355;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Whole;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Wax S.I., Wang J., Heien F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.B.,  
 RA Brownstein M.J., Ueidi T.B., Toshilyki S., Canninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mulhaly S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Wortley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Heltan E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
 RA Krzywinski M.I., Skalska U., Smalins D.E., Schermer A., Schein J.E.,  
 RA Jones S.J., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Whole;  
 RX MEDLINE=22341132; PubMed=12454917;  
 RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,  
 RA Richardson P.;  
 RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus  
 RT Initiative.";  
 RL Dev. Dyn. 225:384-391 (2002).  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Whole;  
 RX Klein S., Strausberg R.;  
 RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: Belongs to class-II aminoacyl-tRNA synthetase family.  
 DR EMBL; BC054256; AAB44256.1; -.  
 DR GO; GO:0005737; C:cytoplasm; IEA.

```

DR GO:0005524; F:ATP binding; IEA.
DR GO:0006486; F:phenylalanine-tRNA ligase activity; IEA.
DR GO:0006432; F:phenylalanine-tRNA aminoacylation; IEA.
DR InterPro: IPR004529; Pfam.
DR InterPro: IPR002319; tRNA-synt_2d.
DR InterPro: IPR006195; tRNA_ligase_11.
DR Pfam: PF01409; tRNA-synt_3d; 1.
DR TIGRFAMs: TIGR00468; pfam; 1.
DR PROSITE: PS50862; AA TRNA_LIGASE_11; 1.
DR SEQUENCE 499 AA; 56750 MW; A632A680B2C27C2 CRC64;

Query Match 100.0%; Score 36; DB 2; Length 499;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 AGPCV 6
Db 121 AGPCV 126

RESULT 6
Q8C7A1 PRELIMINARY; PRT; 798 AA.
ID Q8C7A1;
AC Q8C7A1;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE Mus musculus 13 days embryo heart cDNA, RIKEN full-length enriched
DE library, clone:D330017D22 product:hypothetical protein, full insert
DE sequence.
GN Name=BC025526;
OS Mus musculus (Mouse);
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Heart;
RX MEDLINE=99279253; PubMed=10349636;
RA Carninci P., Hayashizaki Y.;
RT "High-efficiency full-length cDNA cloning.";
RL Mech. Enzymol. 303:19-44(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Heart;
RX MEDLINE=21085660; PubMed=11217851;
RA RIKEN FANTOM Consortium;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Heart;
RA The FANTOM Consortium;
RT "The RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Heart;
RX MEDLINE=2049374; PubMed=11042159;
RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RT "Normalisation and subtraction of cap-trapper-selected cDNAs to
RT prepare full-length cDNA libraries for rapid discovery of new genes.";
RL Genome Res. 10:1617-1630(2000).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Heart;
RX MEDLINE=20530913; PubMed=11076861;
RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,
RA Kono H., Akiyama J., Nishi K., Katsunai T., Tashiro H., Itoh M.,
RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,

```

```

RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kaehiwagi K.,
RA Fujiwaka S., Inoue K., Togawa Y., Izawa M., Ohara E., Wataniki M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsumura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kita A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Heart;
RA Adachi U., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
RA Hoti F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA Kato H., Kawai U., Kojima Y., Kondo S., Kono H., Kouda M., Koya S.,
RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA Nishi K., Nomura K., Numazaki R., Ono M., Onato N., Okazaki Y.,
RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK052272; BAC34911.1; -
DR MGD; MGI:2446190; BC025526.
KW Hypothetical protein.
SQ SEQUENCE 798 AA; 90029 MW; BA33AB859BAFAC CRC64;

Query Match 100.0%; Score 36; DB 2; Length 798;
Best Local Similarity 100.0%; Pred. No. 2.8e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 AGPCV 6
Db 5 AGPCV 10

RESULT 7
Q6G009 PRELIMINARY; PRT; 843 AA.
ID Q6G009;
AC Q6G009;
DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
DE ZAZ0D1 protein.
GN Name=ZAZ0D1;
OS Homo sapiens (Human);
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Bye;
RX MEDLINE=22388257; PubMed=12477932;
RA Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Bueltow K.H., Schaefer J., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heien F.,
RA Diachenko L., Marinsina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares W.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udén T.B., Tothiyuki S., Carninci P., Prange C.,
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
RA Bosak S.A., McGwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Buterfield J.S.,
RA Krzywicki M.I., Skaleja U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.D., Warr M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

```



```
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Eye;
RA Strauberg R.;
RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC072681; AA072681.1; -.
DR InterPro; IPR003323; OTU.
DR InterPro; IPR002653; Znf_A20.
DR Pfam; PF02338; OTU; 1.
DR SMART; SM00259; Znf_A20; 1.
DR PROSITE; PSS0802; OTU; 1.
SQ SEQUENCE 843 AA; 92525 MW; 6D386C864B12EE57 CRC64;

Query Match
Best Local Similarity 100.0%; Score 36; DB 2; Length 843;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGGPCV 6
Db 716 AGGPCV 721

RESULT 8
Q9NO53 PRELIMINARY; PRT; 858 AA.
ID Q9NO53
AC Q9NO53;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE Zinc finger protein Cezanne.
GN Name=CEZANNE;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21356309; PubMed=11463333;
RA Evans P.C., Taylor E.R., Coadwell J., Heynink K., Beyaert R.,
RA Klishaw P.J.;
RL "Isolation and characterization of two novel A20-like proteins.";
DR Biochem. J. 357:617-623(2001).
DR EMBL; AJ293573; CAB97494.1; -.
DR MEROPS; C64.001; -.
DR Genew; HGNC:16683; ZA20D1.
DR InterPro; IPR003323; OTU.
DR InterPro; IPR009060; UBA_1like.
DR Pfam; PF02338; OTU; 1.
DR PROSITE; PSS0802; OTU; 1.
SQ SEQUENCE 858 AA; 94400 MW; A7D6BD280C9387F CRC64;

Query Match
Best Local Similarity 100.0%; Score 36; DB 2; Length 858;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGGPCV 6
Db 731 AGGPCV 736

RESULT 9
Q723K8 PRELIMINARY; PRT; 1639 AA.
ID Q723K8
AC Q723K8;
DT 01-OCT-2003 (TREMBlrel. 25, Created)
DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Hypothetical protein DKFZp686l1653 (Fragment).
GN Name=DKFZp686l1653;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
```

```
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Human endometrium;
RA Wandut R., Heudner D., Mewes H.W., Weill B., Amid C., Oanger A.,
RA Fobo G., Han M., Wiemann S.;
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BX537774; CAD97835.1; -.
DR Hypothetical protein.
FT NON_TER
SQ SEQUENCE 1639 AA; 182478 MW; B58334E1B217B92 CRC64;

Query Match
Best Local Similarity 100.0%; Score 36; DB 2; Length 1639;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGGPCV 6
Db 6 AGGPCV 11

RESULT 10
Q14675 PRELIMINARY; PRT; 1745 AA.
ID Q14675
AC Q14675;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE KIA0169 protein (Fragment).
GN Name=KIA0169;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96281124; PubMed=8724849;
RA Nagase T., Seki N., Ishikawa K., Tanaka A., Nomura N.;
RT "Prediction of the coding sequences of unidentified human genes. V.
RT The coding sequences of 40 new genes (KIA0161-KIA0200) deduced by
RT analysis of cDNA clones from human cell line KG-1.";
RL DNA Res. 3:17-24(1996).
RN [2]
RP SEQUENCE FROM N.A.
RA Ohara O., Nagase T., Kikuno R., Nomura N.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; D79991; BA011486.1; -.
DR Genew; HGNC:17859; NUP188.
FT NON_TER
FT NON_TER
SQ SEQUENCE 1745 AA; 195697 MW; CF74DB49CF6B871 CRC64;

Query Match
Best Local Similarity 100.0%; Score 36; DB 2; Length 1745;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGGPCV 6
Db 1 AGGPCV 6

RESULT 11
Q6ZOH8 PRELIMINARY; PRT; 1761 AA.
ID Q6ZOH8
AC Q6ZOH8;
DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
DE MKIA0169 protein (Fragment).
GN Name=MKIA0169;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
```

```

RN SEQUENCE FROM N.A.
RP TISSUE-Embryonic tail;
RC PubMed=14621295;
RA Okazaki N., Kikuno R., Ohara R., Inamoto S., Koseki H., Hiraoka S.,
RA Suga Y., Nagase T., Ohara O., Koga H.;
RT "Prediction of the coding sequences of mouse homologues of KIA gene:
RT ii. the complete nucleotide sequences of 500 mouse KIAA-homologous
RT cDNAs identified by screening of terminal sequences of cDNA clones
RT randomly sampled from size-fractionated libraries.";
RL DNA Res. 10:167-180(2003).
DR EMBL; AK129073; BAC97883.1; -.
DR InterPro; IPR008938; ARM.
FT NON TER 1
SQ SEQUENCE 1761 AA; 196893 MW; 20CE3DC250EBD93 CRC64;

Query Match 100.0%; Score 36; DB 2; Length 1761;
Best Local Similarity 100.0%; Pred. No. 5.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGRCV 6
Db 7 AGGRCV 12

RESULT 12
BAC97883 PRELIMINARY; PRT; 1761 AA.
AC BAC97883;
DT 02-MAR-2004 (TrEMBLrel. 27, Created)
DT 02-MAR-2004 (TrEMBLrel. 27, Last sequence update)
DT 02-MAR-2004 (TrEMBLrel. 27, Last annotation update)
DE MKIAA0169 protein (Fragment).
GN MKIAA0169.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Embryonic tail;
RA Okazaki N., Kikuno R., Ohara R., Inamoto S., Koseki H., Hiraoka S.,
RA Suga Y., Nagase T., Ohara O., Koga H.;
RT "Prediction of the coding sequences of Mouse Homologues of KIA Gene:
RT ii. The complete nucleotide sequences of 500 Mouse KIAA-homologous
RT cDNAs identified by screening of terminal sequences of cDNA clones
RT randomly sampled from size-fractionated libraries.";
RL DNA Res. 10:167-180(2003).
DR EMBL; AK129073; BAC97883.1; -.
FT NON TER 1
SQ SEQUENCE 1761 AA; 196893 MW; 20CE3DC250EBD93 CRC64;

Query Match 100.0%; Score 36; DB 2; Length 1761;
Best Local Similarity 100.0%; Pred. No. 5.7e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGRCV 6
Db 7 AGGRCV 12

RESULT 13
Q86345 PRELIMINARY; PRT; 324 AA.
AC Q86345;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Putative pectate-lyase (Fragment).
GN Name=upae;
OS Capsicum annuum (Bell pepper).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;

```

```

OC Lamiids; Solanales; Solanaceae; Capsicum.
OX NCBI_TaxID=4072;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=22113994; PubMed=12118879;
RA Marois E., Van den Ackerveken G., Bonas U.;
RT "The xanthomonas type III effector protein AvrBs3 modulates plant gene
RT expression and induces cell hypertrophy in the susceptible host.";
RL Mol. Plant Microbe Interact. 15:637-646(2002).
DR EMBL; AF492632; AAM12784.1; -.
DR GO; GO:0016829; F:lyase activity; IEA.
DR InterPro; IPR002022; Amb_allergen.
DR InterPro; IPR006626; PbH1.
DR InterPro; IPR01050; Pectin_lyas_1like.
DR Pfam; PF00544; Pect_lyase_C_1.
DR PRINTS; PRO0807; AMBALERGEN.
DR SMART; SM00656; Amb_all; 1.
DR SMART; SM00710; PbH1; 2.
KM Lyase.
FT NON TER 324
SQ SEQUENCE 324 AA; 35751 MW; BC2BED36F8B96B7A CRC64;

Query Match 97.2%; Score 35; DB 2; Length 324;
Best Local Similarity 83.3%; Pred. No. 1.9e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGRCV 6
Db 155 AGGRCV 160

RESULT 14
ID Q50225 PRELIMINARY; PRT; 383 AA.
AC Q50225;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE F420-nonreducing hydrogenase (Small subunit (40kDa) of the membrane-
DE bound hydrogenase ii) precursor (EC 1.12.99.-).
GN Name=whc;
OS Methanosarcina mazei (Methanosarcina frisia).
OC Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;
OC Methanosarcinaceae; Methanosarcina.
OX NCBI_TaxID=2209;
RN [1]
RP SEQUENCE FROM N.A.
RC SMPAIN=GO1;
RX MEDLINE=95154297; PubMed=7851393;
RA Deppenmeier U., Blaut M., Lentes S., Herzberg C., Gottschalk G.;
RT "Analysis of the vhoGAC and vhoHAC operons from Methanosarcina mazei
RT strain Go1 both encoding a membrane-bound hydrogenase and a cytochrome
RT b " ;
RL Eur. J. Biochem. 227:261-269(1995).
DR EMBL; X83112; CA58176.1; -.
DR PIR; S67477; S67477.
DR HSSP; P13063; ICCL.
DR GO; GO:0009375; C:hydrogenase complex; IEA.
DR GO; GO:0008901; F:ferredoxin hydrogenase activity; IEA.
DR GO; GO:0005506; F:iron ion binding; IEA.
DR GO; GO:0008137; F:NADH dehydrogenase (ubiquinone) activity; IEA.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR GO; GO:0006120; P:mitochondrial electron transport, NADH to u. . .; IEA.
DR InterPro; IPR006137; Oxidored_q5.
DR InterPro; IPR006311; Tat.
DR Pfam; PF01058; Oxidored_q6; 1.
DR TIGRFAMs; TIGR00391; hycA_1.
DR TIGRFAMs; TIGR01409; Tat_signal_seq; 1.
DR Oxidoreductase; Signal.
KM SIGNAL.
FT CHAIN 1 48 Potential.
FT 49 383 Potential.

```

SQ SEQUENCE 383 AA; 41152 MW; 0F35036868141456 CRC64;  
 Query Match 97.2%; Score 35; DB 2; Length 383;  
 Best Local Similarity 83.3%; Pred. No. 2.2e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 AGGPGCV 6  
 Db 323 AGGPCT 328  
 RESULT 15  
 Q50248 PRELIMINARY; PRT; 383 AA.  
 AC 050248  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)  
 DE F420-nonreducing hydrogenase I precursor.  
 GN Name=vhog;  
 OS Methanosarcina mazel (Methanosarcina frisia).  
 OC Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;  
 OC Methanosarcinaceae; Methanosarcina.  
 OX NCBI\_TaxID=2209;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Goel;  
 RX MEDLINE=95154297; PubMed=7851393;  
 RA Deppenmeier U., Blaut M., Lenters S., Herzberg C., Gottschalk G.;  
 RT "Analysis of the vhogAC and vhtGAC operons from Methanosarcina mazel  
 RT strain Goel both encoding a membrane-bound hydrogenase and a cytochrome  
 RT b";  
 RT Eur. J. Biochem. 227:261-269(1995).  
 DR EMBL; X82940; CA58113.1; -.  
 DR FJ; 549928; S49928.  
 DR HSSP; P13063; ICCL.  
 DR GO; GO:0009375; C:hydrogenase complex; IEA.  
 DR GO; GO:0008901; F:ferredoxin hydrogenase activity; IEA.  
 DR GO; GO:0005506; F:iron ion binding; IEA.  
 DR GO; GO:0008137; F:NADH dehydrogenase (ubiquinone) activity; IEA.  
 DR GO; GO:0006120; P:mitochondrial electron transport, IEA.  
 DR InterPro; IPR001821; Nifendyrog\_small.  
 DR InterPro; IPR006137; Oxidored\_g6.  
 DR InterPro; IPR006311; Tat.  
 DR Pfam; PF01058; Oxidored\_g6\_1.  
 DR TIGRFAMs; TIGR00391; hvdA; 1.  
 DR TIGRFAMs; TIGR01409; TAT\_signal\_seq; 1.  
 KW Signal.  
 FT SIGNAL 1 48 Potential.  
 SQ SEQUENCE 383 AA; 41067 MW; 08C94F527771C9B8 CRC64;  
 Query Match 97.2%; Score 35; DB 2; Length 383;  
 Best Local Similarity 83.3%; Pred. No. 2.2e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 AGGPGCV 6  
 Db 323 AGGPCT 328  
 RESULT 16  
 Q8PUM0 PRELIMINARY; PRT; 386 AA.  
 AC 08PUM0  
 DT 01-OCT-2002 (TREMBlrel. 22, Created)  
 DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)  
 DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)  
 DE F420-nonreducing hydrogenase I (EC 1.1.2.2.-).  
 GN Name=vhog; OrderedLocuNames=MM2314;  
 OS Methanosarcina mazel (Methanosarcina frisia).  
 OC Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;  
 OC Methanosarcinaceae; Methanosarcina.

OX NCBI\_TaxID=2209;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Goel / ATCC BAA-199 / DSM 3647 / OCM 88;  
 RX MEDLINE=22120827; PubMed=12125824;  
 RA Deppenmeier U., Johann A., Hartisch T., Merkl R., Schmitz R.A.,  
 RA Martinez-Arias R., Henne A., Wierer A., Baumeister S., Jacobl C.,  
 RA Brueggemann H., Lienard T., Christmann A., Boemcke M., Steckel S.,  
 RA Bhattacharya A., Lykdis A., Overbeek R., Klenk H.-P., Gunsalus R.P.,  
 RA Fritz H.-J., Gottschalk G.;  
 RT "The genome of Methanosarcina mazel: evidence for lateral gene  
 RT transfer between Bacteria and Archaea."  
 RT J. Mol. Microbiol. Biotechnol. 4:453-461(2002).  
 DR EMBL; AB013474; AMM2010.1; -.  
 DR HSSP; P13063; ICCL.  
 DR GO; GO:0009375; C:hydrogenase complex; IEA.  
 DR GO; GO:0008901; F:ferredoxin hydrogenase activity; IEA.  
 DR GO; GO:0005506; F:iron ion binding; IEA.  
 DR GO; GO:0008137; F:NADH dehydrogenase (ubiquinone) activity; IEA.  
 DR GO; GO:0006118; F:oxidoreductase activity; IEA.  
 DR GO; GO:0006120; P:mitochondrial electron transport, IEA.  
 DR InterPro; IPR006137; Oxidored\_g6.  
 DR InterPro; IPR006311; Tat.  
 DR Pfam; PF01058; Oxidored\_g6\_1.  
 DR TIGRFAMs; TIGR00391; hvdA; 1.  
 DR TIGRFAMs; TIGR01409; TAT\_signal\_seq; 1.  
 KW Complete proteome; Oxidoreductase.  
 SQ SEQUENCE 386 AA; 41427 MW; E80AA4F3B9CEFC7C CRC64;  
 Query Match 97.2%; Score 35; DB 2; Length 386;  
 Best Local Similarity 83.3%; Pred. No. 2.2e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 AGGPGCV 6  
 Db 326 AGGPCT 331  
 RESULT 17  
 Q8PV06 PRELIMINARY; PRT; 386 AA.  
 AC 08PV06  
 DT 01-OCT-2002 (TREMBlrel. 22, Created)  
 DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)  
 DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)  
 DE F420-nonreducing hydrogenase II (EC 1.1.2.2.-).  
 GN Name=vhog; OrderedLocuNames=MM2169;  
 OS Methanosarcina mazel (Methanosarcina frisia).  
 OC Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;  
 OC Methanosarcinaceae; Methanosarcina.  
 OX NCBI\_TaxID=2209;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Goel / ATCC BAA-199 / DSM 3647 / OCM 88;  
 RL J. Mol. Microbiol. Biotechnol. 4:453-461(2002).  
 DR EMBL; AB013457; AMM1865.1; -.  
 DR HSSP; P13063; ICCL.  
 DR GO; GO:0009375; C:hydrogenase complex; IEA.  
 DR GO; GO:0008901; F:ferredoxin hydrogenase activity; IEA.  
 DR GO; GO:0005506; F:iron ion binding; IEA.  
 DR GO; GO:0008137; F:NADH dehydrogenase (ubiquinone) activity; IEA.  
 DR GO; GO:0016491; F:oxidoreductase activity; IEA.  
 DR GO; GO:0006118; P:electron transport; IEA.

DR GO:0006120; P:mitochondrial electron transport, NADH to u. . .; IEA.  
 DR InterPro: IPR001821; NifHydrog\_small.  
 DR InterPro: IPR006137; Oxidored\_g6.  
 DR InterPro: IPR006311; Tat.  
 DR Pfam: PF01058; Oxidored\_g6; 1.  
 DR TIGRFAMs: TIGR00391; HydA\_1.  
 DR TIGRFAMs: TIGR01409; Tat\_signal\_seq; 1.  
 DR Complete proteome; Oxidoreductase.  
 KW SEQUENCE 386 AA; 41511 MW; EPPF68C9A68923C2 CRC64;

Query Match 97.2%; Score 35; DB 2; Length 386;  
 Best Local Similarity 83.3%; Pred. No. 2.2e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGPGCV 6  
 |||||:  
 Db 326 AGGPCT 331

## RESULT 18

Q8A8W8 PRELIMINARY; PRT; 393 AA.  
 AC Q8A8W8:  
 DT 01-JUN-2003 (TRENBLrel. 24, Created)  
 DT 01-JUN-2003 (TRENBLrel. 24, Last sequence update)  
 DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)  
 DE Putative patatin-like protein.  
 GN OrderedLocusNames=BT1049;  
 OS Bacteroides thetaiotaomicron.  
 OC Bacteroides thetaiotaomicron.  
 CC Bacteroidaceae; Bacteroidales; Bacteroides.  
 NCBI\_TaxID=818;

OK NCBI\_TaxID=818;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=VPI-5482 / ATCC 29148;  
 RX MEDLINE=22550858; PubMed=12663928;  
 RA Xu J., Bjursell M.K., Himrod J., Deng S., Carmichael L.K.,  
 RA Chang H.C., Hooper L.V., Gordon J.I.;  
 RT "A genomic view of the human-Bacteroides thetaiotaomicron symbiosis."  
 RL Science 299:2074-2076 (2003).  
 DR EMBL, AB016930; AAO76156.1; -  
 DR InterPro: IPR008985; Cons\_like\_lec\_g1.  
 KW Complete proteome.  
 SQ SEQUENCE 393 AA; 43221 MW; 9158D50A717369A CRC64;

Query Match 97.2%; Score 35; DB 2; Length 393;  
 Best Local Similarity 83.3%; Pred. No. 2.2e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGPGCV 6  
 |||||:  
 Db 293 AGGPCT 298

## RESULT 19

Q9M505 PRELIMINARY; PRT; 398 AA.  
 AC Q9M505:  
 DT 01-OCT-2000 (TRENBLrel. 15, Created)  
 DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)  
 DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)  
 DE Pectate lyase.  
 GN Name=Pel;  
 OS Vitis vinifera (Grape).  
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; rosids;  
 OC Vitaceae; Vitis.  
 OC NCBI\_TaxID=29760;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RA Papadakis G.N., Kanelis A.K.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL, AF243475; AAF63756.1; -

DR HSP; P29155; IJRG.  
 DR GO:0016829; P-lyase activity; IEA.  
 DR InterPro: IPR002022; Amb\_allergen.  
 DR InterPro: IPR006626; Pbl.  
 DR InterPro: IPR011050; Pectin\_lyas\_like.  
 DR Pfam: PF00544; Pec\_lyase\_C; 1.  
 DR PRINTS: PR00807; AMBALLERGEN.  
 DR SMART: SM00656; Amb\_all; 1.  
 DR SMART: SM00710; Pbl; 2.  
 KW Lyase.  
 SQ SEQUENCE 398 AA; 44000 MW; D20A568A1F1241F4 CRC64;

Query Match 97.2%; Score 35; DB 2; Length 398;  
 Best Local Similarity 83.3%; Pred. No. 2.3e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGPGCV 6  
 |||||:  
 Db 151 AGGPCT 156

## RESULT 20

Q93XJ1 PRELIMINARY; PRT; 409 AA.  
 AC Q93XJ1:  
 DT 01-DEC-2001 (TRENBLrel. 19, Created)  
 DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)  
 DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)  
 DE Pectate lyase.  
 GN Name=Sppl;  
 OS Salix gligiana.  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
 OC Euroside 1; Malpighiales; Salicaceae; Salix.  
 NCBI\_TaxID=75556;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Catkin;  
 RA Futamura N., Kouchi H., Shinohara K.;  
 RT "A gene for pectate lyase expressed in elongating and differentiating  
 RT tissues of a Japanese willow (Salix gligiana)."  
 RL J. Plant Physiol. 159:1123-1130(2002).  
 DR EMBL, AB048260; BAB59066.1; -  
 DR HSP; P29155; IJRG.  
 DR GO:0016829; P-lyase activity; IEA.  
 DR InterPro: IPR002022; Amb\_allergen.  
 DR InterPro: IPR006626; Pbl.  
 DR InterPro: IPR011050; Pectin\_lyas\_like.  
 DR Pfam: PF00544; Pec\_lyase\_C; 1.  
 DR PRINTS: PR00807; AMBALLERGEN.  
 DR SMART: SM00656; Amb\_all; 1.  
 DR SMART: SM00710; Pbl; 2.  
 KW Lyase.  
 SQ SEQUENCE 409 AA; 44694 MW; F97E9D0A28CEDFCD CRC64;

Query Match 97.2%; Score 35; DB 2; Length 409;  
 Best Local Similarity 83.3%; Pred. No. 2.3e+02;  
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGPGCV 6  
 |||||:  
 Db 162 AGGPCT 167

## RESULT 21

Q6U7H9 PRELIMINARY; PRT; 418 AA.  
 AC Q6U7H9:  
 DT 05-JUL-2004 (TRENBLrel. 27, Created)  
 DT 05-JUL-2004 (TRENBLrel. 27, Last sequence update)  
 DT 05-JUL-2004 (TRENBLrel. 27, Last annotation update)  
 DE Pectate lyase.  
 OS Malus domestica (Apple) (Malus sylvestris).

```

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eustoids I; Rosales; Rosaceae; Maloideae; Malus.
OX NCBI_TaxID=3750;
RN
  [1]
  SEQUENCE FROM N.A.
  RC TISSUE=Ripe fruit;
  RA Goulae L.F., Oliveira C.M.;
  RL Submitted (Aug-2003) to the EMBL/GenBank/DBJ databases.
  DR EMBL: AY376878; AAC84042.1;
  DR GO: GO:0016829; P:lyase activity; IEA.
  DR InterPro: IPR010502; Amb_allergen.
  DR InterPro: IPR011050; Pectin_lyase_1like.
  DR Pfam: PF00544; Pect_lyase_C_1.
  DR PRINTS: PR00807; AMBALERGEN.
  DR SMART: SM00556; Amb_all; 1.
  KM
  LYASE.
  SQ SEQUENCE 418 AA; 45676 MW; 6FCEA055B929A260 CRC64;

Query Match          97.2%; Score 35; DB 2; Length 418;
Best Local Similarity 83.3%; Pred. No. 2.4e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGFCV 6
Db 171 AGGFCI 176

RESULT 22
AAC84042 PRELIMINARY; PRT; 418 AA.
ID AAC84042;
AC AAC84042;
DT 02-MAR-2004 (TREMBlrel. 27, Last sequence update)
DT 02-MAR-2004 (TREMBlrel. 27, Last annotation update)
DE Pectate lyase.
OS Malus domestica (Apple) (Malus sylvestris).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eustoids I; Rosales; Rosaceae; Maloideae; Malus.
OX NCBI_TaxID=3750;
RN
  [1]
  SEQUENCE FROM N.A.
  RC STRAIN=cv. Royal Gala; TISSUE=Ripe fruit;
  RA Goulae L.F., Oliveira C.M.;
  RL "Isolation and characterization of a pectate lyase (PL) cDNA from ripening apples."
  RT Submitted (Aug-2003) to the EMBL/GenBank/DBJ databases.
  DR EMBL: AY376878; AAC84042.1; -.
  KM
  LYASE.
  SQ SEQUENCE 418 AA; 45676 MW; 6FCEA055B929A260 CRC64;

Query Match          97.2%; Score 35; DB 2; Length 418;
Best Local Similarity 83.3%; Pred. No. 2.4e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGFCV 6
Db 171 AGGFCI 176

RESULT 23
027917 PRELIMINARY; PRT; 501 AA.
ID 027917;
AC 027917;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE Conserved protein.
OS OrderedLocustName=MTN1895;
OS Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteria; Methanobacteriales;
OC Methanobacteriaceae; Methanothermobacter.

```

```

OX NCBI_TaxID=187420;
RN
  [1]
  SEQUENCE FROM N.A.
  RC STRAIN=Delta H;
  RX MEDLINE=96037514; PubMed=9371463;
  RA Smith D.R., Doucette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
  RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
  RA Harrison D., Hoang L., Keagle P., Lumm W., Pochier B., Oiu D.,
  RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
  RA Jivani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
  RA McDougall S., Shimer G., Goyal A., Pietrovski S., Church G.M.,
  RA Daniels C.J., Mao J.-I., Rice P., Noelling J., Reeve J.N.;
  RT "Complete genome sequence of Methanobacterium thermoautotrophicum
  RT delta: functional analysis and comparative genomics."
  RL J. Bacteriol. 179:7135-7155 (1997).
  DR EMBL: AE000941; AAB86355.1; -.
  DR PIR: H69119; H69119.
  DR InterPro: IPR006638; ELP3/MAB/NIFB.
  DR InterPro: IPR007197; Radical SAM.
  DR Pfam: PF04055; Radical SAM; I.
  DR SMART: SM00729; ELP3; I.
  KM
  COMPLETE PROTEOME.
  SQ SEQUENCE 501 AA; 56864 MW; E3319830BF6656C CRC64;

Query Match          97.2%; Score 35; DB 2; Length 501;
Best Local Similarity 83.3%; Pred. No. 2.8e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGFCV 6
Db 94 AGGFCI 99

RESULT 24
ACK1 HUMAN STANDARD; PRT; 1036 AA.
ID ACK1 HUMAN
AC Q07912; Q8N6U7; Q96H59;
DT 29-MAR-2004 (Ref. 43, Created)
DT 29-MAR-2004 (Ref. 43, Last sequence update)
DT 01-OCT-2004 (Ref. 45, Last annotation update)
DE Activated CDC42 kinase 1 (EC 2.7.1.112) (ACK-1).
GN Name=ACK1;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN
  [1]
  SEQUENCE FROM N.A. (ISOFORM 1), AND INTERACTION WITH CDC42.
  RP TISSUE=Hippocampus;
  RC TISSUE=Hippocampus;
  RX MEDLINE=93268389; PubMed=8497321;
  RA Manser E., Leung T., Salimuddin H., Tan L., Lim L.;
  RT "A non-receptor tyrosine kinase that inhibits the GTPase activity of p21cdc42."
  RT Nature 363:364-367 (1993).
  RL
  [2]
  SEQUENCE FROM N.A. (ISOFORM 2).
  RP TISSUE=Brain, and uterus;
  RC TISSUE=Brain, and uterus;
  RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
  RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
  RA Altschul S.F., Zeeberg B., Bueltow K.H., Schaefer C.F., Bhat N.K.,
  RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
  RA Ditschenko L., Marusina K., Farmer A.A., Rubin G.W., Hong L.,
  RA Stapleton M., Soares W.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
  RA Brownstein M.J., Udell T.B., Toshitoki S., Carninci P., Prange C.,
  RA Rana S.S., Loguigliano N.A., Peters G.J., Abramson R.D., Mullany S.J.,
  RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulys S.W.,
  RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
  RA Rane J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
  RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
  RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
  RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,

```



```

RESULT 26
Q8LH84 ID Q8LH84 PRELIMINARY; PRT; 94 AA.
AC Q8LH84;
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
DE Hypothetical protein P0030H06.127 (Hypothetical protein
DE P0475E07.107).
GN Name=P0030H06.127; Synonyms=P0475E07.107;
OS Oryza sativa (Japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RA Sasaki T., Matsumoto T., Yamamoto K.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Sasaki T., Matsumoto T., Yamamoto K.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP004395; BAC10207.1; -
DR EMBL; AP004568; BAC16099.1; -
KW Gramene; Q8LH84; -
KW Hypothetical protein.
SQ SEQUENCE 94 AA; 10446 MW; 86E47D0F6283429D CRC64;

Query Match 91.7%; Score 33; DB 1; Length 94;
Best Local Similarity 83.3%; Pred. No. 7.6e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGGPGCV 6
Db 69 AGGPCL 74

RESULT 27
UL49 EBV ID UL49 EBV STANDARD; PRT; 591 AA.
AC P14347;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Protein BFRF2.
GN Name=BFRF2;
OS Epstein-Barr virus (strain B95-8) (HHV-4) (Human herpesvirus 4).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Lymphocryptovirus.
OX NCBI_TaxID=10377;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=84270667; PubMed=6087149;
RX Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,
RA Gibson J.J., Hatfull G., Hudson G.S., Satchwell S.C., Seguin C.,
RA Tufnell P.S., Barrell B.G.;
RT "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";
RL Nature 310:207-211(1984).
RN [5]
RP SEQUENCE FROM N.A.
RA MEDLINE=83109311; PubMed=6296170;
RX Deininger P.L., Bankier A., Farrell P., Baer R., Barrell B.;
RT "Sequence analysis and in vitro transcription of portions of the
RT Epstein-Barr virus genome.";
RL J. Cell. Biochem. 19:267-274(1982).
RN [6]
RP SEQUENCE FROM N.A.
RA MEDLINE=83169725; PubMed=6300857;
RX Farrell P.J., Deininger P.L., Bankier A., Barrell B.;
RT "Homologous upstream sequences near Epstein-Barr virus promoters.";

```

```

KW Early protein.
SQ SEQUENCE 591 AA; 63977 MW; BE63FF1E0721912E CRC64;

Query Match 91.7%; Score 33; DB 1; Length 591;
Best Local Similarity 83.3%; Pred. No. 7.6e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGGPGCV 6
Db 333 AGGPCL 338

RESULT 28
Q777G6 ID Q777G6 PRELIMINARY; PRT; 591 AA.
AC Q777G6;
DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
DE BFRF2 protein.
GN Name=BFRF2;
OS Human herpesvirus 4 (Epstein-Barr virus).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Lymphocryptovirus.
OX NCBI_TaxID=10376;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=84270667; PubMed=6087149;
RX Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,
RA Gibson J.J., Hatfull G.F., Hudson G.S., Satchwell S.C., Seguin C.,
RA Tufnell P.S., Barrell B.G.;
RT "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";
RL Nature 310:207-211(1984).
RN [2]
RP SEQUENCE FROM N.A.
RA MEDLINE=84270667; PubMed=6087149;
RX Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,
RA Gibson J.J., Hatfull G.F., Hudson G.S., Satchwell S.C., Seguin C.,
RA Tufnell P.S., Barrell B.G.;
RT "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";
RL Nature 310:207-211(1984).
RN [3]
RP SEQUENCE FROM N.A.
RA MEDLINE=82014887; PubMed=6269068;
RX Arrand J.R., Rymo L., Walsh J.E., Bjorck E., Lindahl T., Griffin B.E.;
RT "Molecular cloning of the complete Epstein-Barr virus genome as a set
RT of overlapping restriction endonuclease fragments.";
RL Nucleic Acids Res. 9:299-3014(1981).
RN [4]
RP SEQUENCE FROM N.A.
RA MEDLINE=82059504; PubMed=7301588;
RX Korak M.;
RT "Possible role of flanking nucleotides in recognition of the AUG
RT initiator codon by eukaryotic ribosomes.";
RL Nucleic Acids Res. 9:5233-5252(1981).
RN [5]
RP SEQUENCE FROM N.A.
RA MEDLINE=83109311; PubMed=6296170;
RX Deininger P.L., Bankier A., Farrell P., Baer R., Barrell B.;
RT "Sequence analysis and in vitro transcription of portions of the
RT Epstein-Barr virus genome.";
RL J. Cell. Biochem. 19:267-274(1982).
RN [6]
RP SEQUENCE FROM N.A.
RA MEDLINE=83169725; PubMed=6300857;
RX Farrell P.J., Deininger P.L., Bankier A., Barrell B.;
RT "Homologous upstream sequences near Epstein-Barr virus promoters.";

```

Proc. Natl. Acad. Sci. U.S.A. 80:1565-1569(1983).

[17] SEQUENCE FROM N.A.

RA STRAIN=B95-8;

RC MEDLINE=85035713; PubMed=6092825;

RA Bankier A.T., Deininger P.L., Farrell P.J., Barrell B.G.;

RT "Sequence analysis of the 17,166 base-pair EcoRI fragment C of B95-8

RT Epstein-Barr virus.";

RL Mol. Biol. Med. 1:21-45(1983).

[18] SEQUENCE FROM N.A.

RA STRAIN=B95-8;

RC MEDLINE=85060424;

RA Segun C., Farrell P.J., Barrell B.G.;

RT "DNA sequence and transcription of the BamHI fragment B region of B95-

RT 8 Epstein-Barr virus.";

RL Mol. Biol. Med. 1:369-392(1983).

[19] SEQUENCE FROM N.A.

RA STRAIN=B95-8;

RC MEDLINE=83294686; PubMed=6310141;

RA Jeang K.T., Hayward S.D.;

RT "Organization of the Epstein-Barr virus DNA molecule. III. Location of

RT the P3HR-1 deletion junction and characterization of the NotI repeat

RT units that form part of the template for an abundant 12-O-

RT tetradecanoylphorbol-13-acetate-induced mRNA transcript.";

RL J. Virol. 48:135-148(1983).

[10] SEQUENCE FROM N.A.

RA STRAIN=B95-8;

RC MEDLINE=85060428; PubMed=6094955;

RA Bankier A.T., Deininger P.L., Satchwell S.C., Baer R., Farrell P.J.,

RA Barrell B.G.;

RT "DNA sequence analysis of the EcoRI Dhet fragment of B95-8 Epstein-

RT Barr virus containing the terminal repeat sequences.";

RL Mol. Biol. Med. 1:425-445(1983).

[11] SEQUENCE FROM N.A.

RA STRAIN=B95-8;

RC MEDLINE=20331131; PubMed=10872327;

RA Farrell P.J., Bankier A., Segun C., Deininger P., Barrell B.G.;

RT "Latent and lytic cycle promoters of Epstein-Barr virus.";

RL EMBO J. 2:1331-1338(1983).

[12] SEQUENCE FROM N.A.

RA STRAIN=B95-8;

RC MEDLINE=84207939; PubMed=6327290;

RA Jones M.D., Foster L., Sheedy T., Griffin B.E.;

RT "The EB virus genome in David Burkitt's lymphoma cells has a deletion

RT similar to that observed in a non-transforming strain (P3HR-1) of the

RT virus.";

RL EMBO J. 3:813-821(1984).

[13] SEQUENCE FROM N.A.

RA STRAIN=B95-8;

RC MEDLINE=84236104; PubMed=6203743;

RA Biggin M., Farrell P.J., Barrell B.G.;

RT "Transcription and DNA sequence of the BamHI L fragment of B95-8

RT Epstein-Barr virus.";

RL EMBO J. 3:1083-1090(1984).

[14] SEQUENCE FROM N.A.

RA STRAIN=B95-8;

RC MEDLINE=8422045; PubMed=6328526;

RA Yates J., Warren N., Reisman D., Sugden B.;

RT "A cis-acting element from the Epstein-Barr viral genome that permits

RT stable replication of recombinant plasmids in latently infected

RT cells.";

RL Proc. Natl. Acad. Sci. U.S.A. 81:3806-3810(1984).

[15] SEQUENCE FROM N.A.

RA STRAIN=B95-8;

RC MEDLINE=84247360; PubMed=6330697;

RA Gibson T., Stockwell P., Ginsburg M., Barrell B.;

RT "Homology between two EBV early genes and HSV ribonucleotide reductase

RT and 38 genes.";

RL Nucleic Acids Res. 12:5087-5099(1984).

[16] SEQUENCE FROM N.A.

RA STRAIN=B95-8;

RC MEDLINE=87289053; PubMed=3039467;

RA Boddecoot M., Pericandet M.;

RT "Clustered alternative splice sites in Epstein-Barr virus RNAs.";

RL Nucleic Acids Res. 15:5887-5887(1987).

[17] SEQUENCE FROM N.A.

RA STRAIN=B95-8;

RC MEDLINE=91021036; PubMed=2171209;

RA Parker B.D., Bankier A., Satchwell S., Barrell B., Farrell P.J.;

RT "Sequence and transcription of Raji Epstein-Barr virus DNA spanning

RT the B95-8 deletion region.";

RL Virology 179:339-346(1990).

[18] SEQUENCE FROM N.A.

RA STRAIN=B95-8;

RC Hattful G.F., Barrell B.G., Quinn J., McGeoch D.;

RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.

[19] SEQUENCE FROM N.A.

RA STRAIN=B95-8;

RC Blum U.K., Amon W., Farrell P.J.;

RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; AJ507799; CAD53400.1; -

DR InterPro; IPR004339; UL49.

DR Pfam; PFO3117; Herpes\_UL49\_1; 1.

SO SEQUENCE 591 AA; 63977 MW; EE63FF1E0721912E CRC64;

Query Match 91.7%; Score 33; DB 2; Length 591;

Best Local Similarity 83.3%; Pred. NO. 7.6e+02;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGPCV 6

Db 333 AGPCV 338

RESULT 29

ID CAD53400 PRELIMINARY; PRT; 591 AA.

AC CAD53400:

DT 10-MAY-2004 (TrEMBLrel. 27, Created)

DT 10-MAY-2004 (TrEMBLrel. 27, Last sequence update)

DT 10-MAY-2004 (TrEMBLrel. 27, Last annotation update)

DE BFRF2 protein.

DE BFRF2.

GN Human herpesvirus 4 (Epstein-Barr virus).

OS Viruses; dsDNA viruses, no RNA stage; Herpesviridae;

OC Gammaherpesvirinae; Lymphocryptovirus.

OX NCBI\_TaxID=10376;

[1] SEQUENCE FROM N.A.

RA STRAIN=B95-8;

RC MEDLINE=82014687; PubMed=6269068;

RA Arrand J.R., Rymo L., Walsh J.E., Bjorck E., Lindahl T., Griffin B.E.;

RT "Molecular cloning of the complete Epstein-Barr virus genome as a set

RT of overlapping restriction endonuclease fragments.";

RL Nucleic Acids Res. 9:2999-3014(1981).

[2] SEQUENCE FROM N.A.

RA STRAIN=B95-8;

RC MEDLINE=82059504; PubMed=7301588;

RA Kozak M.;

RT "Possible role of flanking nucleotides in recognition of the AUG

RT initiator codon by eukaryotic ribosomes.";

RL Nucleic Acids Res. 9:5233-5252(1981).

[3]



RP SEQUENCE FROM N.A.  
RC STRAIN=B95-8;  
RX MEDLINE=83109311; PubMed=6296170;  
RA Deininger P.L., Bankier A., Farrell P., Baer R., Barrell B.;  
RT "Sequence analysis and in vitro transcription of portions of the  
RL Epstein-Barr virus genome.";  
RN J. Cell. Biochem. 19:267-274(1982).  
[4]  
RP SEQUENCE FROM N.A.  
RC STRAIN=B95-8;  
RX MEDLINE=83169725; PubMed=6300857;  
RA Farrell P.J., Deininger P.L., Bankier A., Barrell B.;  
RT "Homologous upstream sequences near Epstein-Barr virus promoters.";  
RL Proc. Natl. Acad. Sci. U.S.A. 80:1565-1569(1983).  
[5]  
RP SEQUENCE FROM N.A.  
RC STRAIN=B95-8;  
RX MEDLINE=85035713; PubMed=6092825;  
RA Bankier A.T., Deininger P.L., Farrell P.J., Barrell B.G.;  
RT "Sequence analysis of the 17,166 base-pair EcoRI fragment C of B95-8  
RL Epstein-Barr virus.";  
RN Mol. Biol. Med. 1:21-45(1983).  
[6]  
RP SEQUENCE FROM N.A.  
RC STRAIN=B95-8;  
RX MEDLINE=85060424; PubMed=6094953;  
RA Seguin C., Farrell P.J., Barrell B.G.;  
RT "DNA sequence and transcription of the BamHI fragment B region of B95-  
RL 8 Epstein-Barr virus.";  
RN Mol. Biol. Med. 1:369-392(1983).  
[7]  
RP SEQUENCE FROM N.A.  
RC STRAIN=B95-8;  
RX MEDLINE=83294686; PubMed=6310141;  
RA Jeang K.T., Hayward S.D.;  
RT "Organization of the Epstein-Barr virus DNA molecule. III. Location of  
RT the P3HR-1 deletion junction and characterization of the NotI repeat  
RT units that form part of the template for an abundant 12-O-  
RL tetradeconoylphorbol-13-acetate-induced mRNA transcript.";  
RN J. Virol. 48:1135-1148(1983).  
[8]  
RP SEQUENCE FROM N.A.  
RC STRAIN=B95-8;  
RX MEDLINE=85060428; PubMed=6094955;  
RA Bankier A.T., Deininger P.L., Satchwell S.C., Baer R., Farrell P.J.,  
RT "DNA sequence analysis of the EcoRI Dhet fragment of B95-8 Epstein-  
RL Barr virus containing the terminal repeat sequences.";  
RN Mol. Biol. Med. 1:425-445(1983).  
[9]  
RP SEQUENCE FROM N.A.  
RC STRAIN=B95-8;  
RX MEDLINE=20331131; PubMed=10872327;  
RA Farrell P.J., Bankier A., Seguin C., Deininger P., Barrell B.G.;  
RT "Latent and lytic cycle promoters of Epstein-Barr virus.";  
RN EMBO J. 2:1331-1338(1983).  
[10]  
RP SEQUENCE FROM N.A.  
RC STRAIN=B95-8;  
RX MEDLINE=84207939; PubMed=6327290;  
RA Jones M.D., Foster L., Sheedy T., Griffin B.E.;  
RT "The EB virus genome in Daudi Burkitt's lymphoma cells has a deletion  
RL similar to that observed in a non-transforming strain (P3HR-1) of the  
RT virus.";  
RN EMBO J. 3:813-821(1984).  
[11]  
RP SEQUENCE FROM N.A.  
RC STRAIN=B95-8;  
RX MEDLINE=84236104; PubMed=6203743;  
RA Biggin M., Farrell P.J., Barrell B.G.;  
RT "Transcription and DNA sequence of the BamHI L fragment of B95-8  
RL Epstein-Barr virus.";  
RN EMBO J. 3:1083-1090(1984).

RN [12]  
RP SEQUENCE FROM N.A.  
RC STRAIN=B95-8;  
RX MEDLINE=84222045; PubMed=6328526;  
RA Yates J., Warren N., Reisman D., Sugden B.;  
RT "A cis-acting element from the Epstein-Barr viral genome that permits  
RT stable replication of recombinant plasmids in latently infected  
RT cells.";  
RL Proc. Natl. Acad. Sci. U.S.A. 81:3806-3810(1984).  
[13]  
RP SEQUENCE FROM N.A.  
RC STRAIN=B95-8;  
RX MEDLINE=84247360; PubMed=6330697;  
RA Gibson T., Stockwell P., Ginsburg M., Barrell B.;  
RT "Homology between two EBV early genes and HSV ribonucleotide reductase  
RT and 38k genes.";  
RN Nucleic Acids Res. 12:5087-5099(1984).  
[14]  
RP SEQUENCE FROM N.A.  
RC STRAIN=B95-8;  
RX MEDLINE=84270667; PubMed=6087149;  
RA Baer R.J., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,  
RT Gibson T.J., Hatfull G.F., Hudson G.S., Satchwell S.C., Seguin C.,  
RL Tuffnell P.S., Barrell B.G.;  
RT "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";  
RN Nature 310:207-211(1984).  
[15]  
RP SEQUENCE FROM N.A.  
RC STRAIN=B95-8;  
RX MEDLINE=87289053; PubMed=3039467;  
RA Bodescot M., Perricaudet M.;  
RT "Clustered alternative splice sites in Epstein-Barr virus RNAs.";  
RL Nucleic Acids Res. 15:5887-5887(1987).  
[16]  
RP SEQUENCE FROM N.A.  
RC STRAIN=B95-8;  
RX MEDLINE=88283646; PubMed=2840285;  
RA Laux G., Perricaudet M., Farrell P.J.;  
RT "A spliced Epstein-Barr virus gene expressed in immortalized  
RT lymphocytes is created by circularization of the linear viral  
RT genome.";  
RN EMBO J. 7:769-774(1988).  
[17]  
RP SEQUENCE FROM N.A.  
RC STRAIN=B95-8;  
RX MEDLINE=91021036; PubMed=2171209;  
RA Parker B.D., Bankier A., Satchwell S., Barrell B., Farrell P.J.;  
RT "Sequence and transcription of Raji Epstein-Barr virus DNA spanning  
RL the B95-8 deletion region.";  
RN Virology 179:339-346(1990).  
[18]  
RP SEQUENCE FROM N.A.  
RC STRAIN=B95-8;  
RX Hatfull G.F., Barrell B.G., Quinn J., McGeoch D.;  
RL Submitted (OCT-2002) to the EMBL/GenBank/DBJ databases.  
[19]  
RP SEQUENCE FROM N.A.  
RC STRAIN=B95-8;  
RX Blime U.K., Amon W., Farrell P.J.;  
RT "Induction of Epstein-Barr virus late promoters on small plasmids in  
RL the EBV late lytic cycle requires ori lyt.";  
RN Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.  
[20]  
RP SEQUENCE FROM N.A.  
RC STRAIN=B95-8;  
RX Farrell P.J.;  
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AJ507799; CAD53400.1; -;  
SQ SEQUENCE 591 AA; 63977 MW; EE63FF1E0721912B CRC64;  
Query Match 91.7%; Score 33; DB 2; Length 591;  
Best Local Similarity 83.3%; Pred. No. 7.6e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGPCV 6  
|||||  
DB 333 AGPCV 338

## RESULT 30

06RKJ9 PRELIMINARY; PRT: 2434 AA.  
AC 06RKJ9  
DT 05-JUL-2004 (TREMBlrel. 27, Created)  
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)  
DE Polyketide synthase (Fragment).  
CN Name=PKS1;  
OS Botrytis cinerea (Noble rot fungus) (Botryotinia fuckeliana).  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Leotiomycetes;  
OC Helotiales; Sclerotiniaceae; Botryotinia.  
OX NCBI\_TaxID=40559;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=B05.10;  
RX PubMed=14676319;  
RA Krokken S., Glass N.L., Taylor J.W., Yoder O.C., Turgeon B.G.;  
RT "Phylogenomic analysis of type I polyketide synthase genes in  
pathogenic and saprobic ascomycetes.";  
RT Proc. Natl. Acad. Sci. U.S.A. 100:15670-15675(2003).  
RL EMBL: AY495606; AAR90237.1; -  
DR InterPro: IPR009081; ACP-like.  
DR InterPro: IPR001227; ACP\_transferase.  
DR InterPro: IPR002085; Adh\_zn\_family.  
DR InterPro: IPR011032; GroES-like.  
DR InterPro: IPR000794; Ketoacyl\_synth.  
DR InterPro: IPR006163; Pp\_bind.  
DR InterPro: IPR000051; SAM\_bind.  
DR Pfam: PF00698; ACP\_transferase\_1; 1.  
DR Pfam: PF00107; ADH\_zinc\_N; 1.  
DR Pfam: PF00109; ketoacyl-synth\_C; 1.  
DR Pfam: PF02801; ketoacyl-synth\_C; 1.  
DR PROSITE: PS50075; ACP\_DOMAIN; 1.  
DR PROSITE: PS00606; B\_KETOACYL\_SYNTHASE; 1.  
DR TRANSFERASE.  
FT NON\_TER 1  
FT TER 2434 2434  
SQ SEQUENCE 2434 AA; 269939 MW; 55887CFDC59CE245 CRC64;  
Query Match 91.7%; Score 33; DB 2; Length 2434;  
Best Local Similarity 83.3%; Pred. No. 2.7e+03;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGPCV 6  
|||||  
DB 1511 SGPCV 1516

## RESULT 31

AAR90237 PRELIMINARY; PRT: 2434 AA.  
AC AAR90237;  
DT 02-MAR-2004 (TREMBlrel. 27, Created)  
DT 02-MAR-2004 (TREMBlrel. 27, Last sequence update)  
DE Polyketide synthase (Fragment).  
CN PKS1.  
OS Botrytis cinerea (Noble rot fungus) (Botryotinia fuckeliana).  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Leotiomycetes;  
OC Helotiales; Sclerotiniaceae; Botryotinia.  
OX NCBI\_TaxID=40559;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=B05.10;  
RX PubMed=14676319;  
RA Krokken S., Glass N.L., Taylor J.W., Yoder O.C., Turgeon B.G.;

RT "Phylogenomic analysis of type I polyketide synthase genes in  
pathogenic and saprobic ascomycetes.";  
RT Proc. Natl. Acad. Sci. U.S.A. 100:15670-15675(2003).  
RL EMBL: AY495606; AAR90237.1; -  
FT NON\_TER 1  
FT TER 2434 2434  
SQ SEQUENCE 2434 AA; 269939 MW; 55887CFDC59CE245 CRC64;

OY 1 AGPCV 6  
|||||  
DB 1511 SGPCV 1516

## RESULT 32

066009 PRELIMINARY; PRT: 73 AA.  
AC 066009;  
DT 01-AUG-1998 (TREMBlrel. 07, Created)  
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)  
DE Phosphofructokinase I (EC 2.7.1.11) (Fragment).  
CN Name=pfkI;  
OS Bacillus subtilis.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=1423;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 6057;  
RX MEDLINE=20388102; PubMed=10932722;  
RA Munoz M.E., Leborgne S., Bolivar F., Valle F.;  
RT "Molecular cloning of the gene that codes for the pyruvate kinase of  
Bacillus subtilis, primary characterization of a strain carrying this  
gene insertionally inactivated.";  
RT Rev. Latinoam. Microbiol. 39:129-140(1997).  
RL EMBL: U73943; AAC12961.1; -  
DR HSPF: P00512; 3PFK.  
DR GO: GO:0005945; C:6-phosphofructokinase complex; IEA.  
DR GO: GO:0003872; F:6-phosphofructokinase activity; IEA.  
DR GO: GO:0016301; F:kinase activity; IEA.  
DR GO: GO:0016740; F:transferase activity; IEA.  
DR GO: GO:0006096; P:glycolysis; IEA.  
DR InterPro: IPR000023; Pfrfructinase.  
DR Pfam: PF00365; PFK; 1.  
DR KINASE; Transferase.  
FT NON\_TER 1  
FT TER 73 73  
SQ SEQUENCE 73 AA; 7969 MW; 4768F4F9DB420AB8 CRC64;  
Query Match 88.9%; Score 32; DB 2; Length 73;  
Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GGPCV 6  
|||||  
DB 34 GGPCV 38

## RESULT 33

075WH6 PRELIMINARY; PRT: 77 AA.  
AC 075WH6;  
DT 05-JUL-2004 (TREMBlrel. 27, Created)  
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)  
DE Peptide toxin 1.  
OS Macrochele gigas (Spider).  
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;  
OC Mygalomorphae; Hexatheleidae; Macrochele.  
OX NCBI\_TaxID=223896;  
RN [1]

RP SEQUENCE FROM N.A.  
 RC TISSUE=Venom gland;  
 RA Satake H., Vallegas E., Corzo G.;  
 RL Submitted (SEP-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB121195; BAD13402.1;  
 SQ SEQUENCE 77 AA; 8255 MW; B304C320CC9E1318 CRC64;

Query Match 88.9%; Score 32; DB 2; Length 77;  
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GPCV 6  
 |||||  
 DB 50 GPCV 54

## RESULT 34

BAD13402 PRELIMINARY; PRT; 77 AA.  
 AC BAD13402;  
 DT 31-MAR-2004 (TRENBLREL. 27, Created)  
 DT 31-MAR-2004 (TRENBLREL. 27, Last sequence update)  
 DT 31-MAR-2004 (TRENBLREL. 27, Last annotation update)  
 DE Peptide toxin 1.  
 OS Macrochele gigas (Spider).  
 CC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;  
 OC Mygalomorphae; Hexachelidae; Macrochele.  
 CX NCBI\_TaxID=223896;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Venom gland;  
 RA Satake H., Vallegas E., Corzo G.;  
 RL "cDNA encoding a peptide toxin of Macrochele gigas";  
 DR EMBL; AB121195; BAD13402.1;  
 SQ SEQUENCE 77 AA; 8255 MW; B304C320CC9E1318 CRC64;

Query Match 88.9%; Score 32; DB 2; Length 77;  
 Best Local Similarity 100.0%; Pred. No. 1.8e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GPCV 6  
 |||||  
 DB 50 GPCV 54

## RESULT 35

Q7UYG4 PRELIMINARY; PRT; 91 AA.  
 AC Q7UYG4;  
 DT 01-OCT-2003 (TRENBLREL. 25, Created)  
 DT 01-OCT-2003 (TRENBLREL. 25, Last sequence update)  
 DT 01-OCT-2003 (TRENBLREL. 25, Last annotation update)  
 DE Hypothetical protein.  
 GN OrderedLocusNames=RB629;  
 OS Rhodopirella baltica.  
 CC Bacteria; Planctomycetes; Planctomycetacia; Planctomycetales;  
 OC Planctomycetaceae; Pirellula.  
 CX NCBI\_TaxID=117;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=1;  
 RC MEDLINE=22735913; PubMed=12835416;  
 RA Gloeckner F.O., Kube M., Bauer M., Teeling H., Lombardot T.,  
 RA Ludwig W., Gade D., Beck A., Borzym K., Heilmann K., Rabus R.,  
 RA Schleuter H., Amann R., Reinhardt R.;  
 RL "Complete genome sequence of the marine planctomycete Pirellula sp.  
 strain 1";  
 RT Proc. Natl. Acad. Sci. U.S.A. 100:8298-8303(2003).  
 DR EMBL; BX294134; CAD71678.1;  
 KW Complete proteome; Hypothetical protein.  
 SQ SEQUENCE 91 AA; 9627 MW; 6FE56CA3D69FBF3B CRC64;

Query Match 88.9%; Score 32; DB 2; Length 91;  
 Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GPCV 6  
 |||||  
 DB 78 GPCV 82

## RESULT 36

Q9CTQ0 PRELIMINARY; PRT; 106 AA.  
 AC Q9CTQ0;  
 DT 01-JUN-2001 (TRENBLREL. 17, Created)  
 DT 01-JUN-2001 (TRENBLREL. 17, Last sequence update)  
 DT 01-OCT-2003 (TRENBLREL. 25, Last annotation update)  
 DE Mus musculus adult retina cDNA, RIKEN full-length enriched library,  
 DE clone:A930010C08 product:hypothetical protein, full insert sequence.  
 DE (Fragment).  
 GN Name=Ntrg1;  
 OS Mus musculus (Mouse).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 CX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Retina;  
 RX MEDLINE=99279253; PubMed=10349636;  
 RA Carninci P., Hayashizaki Y.;  
 RL "High-efficiency full-length cDNA cloning";  
 Meth. Enzymol. 303:19-44(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Retina;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA RIKEN FANTOM Consortium;  
 RL "Functional annotation of a full-length mouse cDNA collection";  
 Nature 409:685-690(2001).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Retina;  
 RA The FANTOM Consortium;  
 RL "Analysis of the mouse transcriptome based on functional annotation of  
 60,770 full-length cDNAs";  
 Nature 420:563-573(2002).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Retina;  
 RX MEDLINE=20499374; PubMed=11042159;  
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,  
 RA Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.;  
 RL "Normalization and subtractions of cap-trapper-selected cDNAs to  
 prepare full-length cDNA libraries for rapid discovery of new genes";  
 Genome Res. 10:1617-1630(2000).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Retina;  
 RX MEDLINE=20530913; PubMed=11076861;  
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,  
 RA Kono H., Akiyama J., Nishi K., Kitsuwa T., Tashiro H., Itoh M.,  
 RA Suni N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,  
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,  
 RA Fujiwara K., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,  
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,  
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
 RL "RIKEN integrated sequence analysis (RISA) system-384-format  
 sequencing pipeline with 384 multicapillary sequencer";  
 Genome Res. 10:1757-1771(2000).  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Retina;  
 RA Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,

```

RA Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,
RA Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F.,
RA Imotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,
RA Kawai J., Kojima Y., Kono H., Kouda M., Koya S., Kurihara C.,
RA Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,
RA Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K.,
RA Sano H., Sasaki D., Shibata K., Shibata Y., Shinagawa A., Shiraki T.,
RA Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,
RA Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,
RA Muramatsu M., Hayashizaki Y.,
RA Submitted (Aug-2000) to the EMBL/GenBank/DBJ databases.
RL EMBL; AK020842; BAB32225.1; -.
DR MGD; MGI:1934028; Ntng1.
DR GO; GO:0005886; C:plasma membrane; IDA.
DR GO; GO:0007409; P:axonogenesis; IDA.
KM Hypothetical protein.
FT NON_TER
SQ SEQUENCE 106 AA; 10899 MW; 575533A93108EBE7 CRC64;
Query Match 88.9%; Score 32; DB 2; Length 106;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 AGPC 5
Db 16 AGPC 20
RESULT 37
Q8BPN6 PRELIMINARY; PRT; 108 AA.
ID Q8BPN6;
AC Q8BPN6;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Mus musculus 0 day neonate eyeball cDNA, RIKEN full-length enriched
DE library, clone:EI30203P04 product:hypothetical protein, full insert
DE sequence. (Fragment).
GN Name=Ntng1;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
RN RP STRAIN=C57BL/6J; TISSUE=Eyeball;
RC MEDLINE=99279253; PubMed=10349636;
RX Carninci P., Hayashizaki Y.;
RA "High-efficiency full-length cDNA cloning.";
RL Mech. Enzymol. 303:19-44(1999).
[2]
RN RP STRAIN=C57BL/6J; TISSUE=Eyeball;
RC MEDLINE=21085660; PubMed=11217851;
RX RIKEN FANTOM Consortium;
RA "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
[3]
RN RP STRAIN=C57BL/6J; TISSUE=Eyeball;
RC The FANTOM Consortium;
RA The FANTOM Consortium;
RA "Analysis of the mouse transcriptome based on functional annotation of
RA 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
[4]
RN RP STRAIN=C57BL/6J; TISSUE=Eyeball;
RC MEDLINE=20499374; PubMed=11042159;
RX Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,
RA Kono H., Okazaki Y., Muramatsu M., Hayashizaki Y.;
RA "Normalization and subtraction of cap-trapper-selected cDNAs to
RA prepare full-length cDNA libraries for rapid discovery of new genes.";
RT

```

```

RL Genome Res. 10:1617-1630(2000).
[5]
RN RP STRAIN=C57BL/6J; TISSUE=Eyeball;
RC MEDLINE=20530913; PubMed=11076861;
RX Shibata K., Itoh M., Aizawa K., Nagao S., Sasaki N., Carninci P.,
RA Kono H., Akiyama Y., Nishi K., Kitaura T., Tashiro H., Itoh M.,
RA Suniti N., Ishii Y., Nakamura S., Hazama M., Nishino T., Harada A.,
RA Yamamoto R., Matsunoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,
RA Fujiwara S., Inoue K., Togawa Y., Izawa M., Obara E., Watanagi M.,
RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsura S., Kawai J.,
RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;
RT "RIKEN integrated sequence analysis (RISA) system-384-format
RT sequencing pipeline with 384 multicapillary sequencer.";
RL Genome Res. 10:1757-1771(2000).
[6]
RN RP STRAIN=C57BL/6J; TISSUE=Eyeball;
RC Adachi J., Aizawa K., Akimura T., Arakawa T., Bono H., Carninci P.,
RA Fukuda S., Furuno M., Hanagaki T., Hara A., Hashizume W.,
RA Hayashida K., Hayatsu N., Hiramoto K., Hiraoka T., Hirozane T.,
RA Hori F., Imotani K., Ishii Y., Itoh M., Kagawa I., Kasukawa T.,
RA Katoh H., Kawai J., Kojima Y., Kondo S., Kono H., Kouda M., Koya S.,
RA Kurihara C., Matsuyama T., Miyazaki A., Murata M., Nakamura M.,
RA Nishi K., Nomura K., Numazaki R., Ohno M., Ohnato N., Okazaki Y.,
RA Saito R., Saitoh H., Sakai C., Sakai K., Sakazume N., Sano H.,
RA Sasaki D., Shibata K., Shinagawa A., Shiraki T., Sogabe Y., Tagami M.,
RA Tagawa A., Takahashi F., Takaku-Akahira S., Takeda Y., Tanaka T.,
RA Tomaru A., Toya T., Yasunishi A., Muramatsu M., Hayashizaki Y.;
RL Submitted (Jul-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK053689; BAC35475.1; -.
DR MGD; MGI:1934028; Ntng1.
DR GO; GO:0005886; C:plasma membrane; IDA.
DR GO; GO:0007409; P:axonogenesis; IDA.
KM Hypothetical protein.
FT NON_TER
SQ SEQUENCE 108 AA; 11167 MW; 186CA944B59AEFA CRC64;
Query Match 88.9%; Score 32; DB 2; Length 108;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 AGPC 5
Db 18 AGPC 22
RESULT 38
Q7KUM2 PRELIMINARY; PRT; 119 AA.
ID Q7KUM2;
AC Q7KUM2;
DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
DE CG33259-PA.
GN ORFNames=CG33259;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
[1]
RN RP STRAIN=C57BL/6J; TISSUE=Eyeball;
RC MEDLINE=20196006; PubMed=10731132;
RX Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter B.G., Helt G., Nelson C.R., Gaber G.L.,
RA Abriil J.F., Agbayani A., An H.U., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,

```

RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
 RA Buttle K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrera S., Fleischmann W.,  
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Gloder A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,  
 RA Jalaal M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Pacle J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.Y., Waasman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodger, Morley K.C., Wu D., Yang S., Yao Q.A., Ye J.,  
 RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu H.O.,  
 RA Zibbe R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195 (2000).  
 [2]  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE=22426065; PubMed=12537568;  
 RA Celniker S.E., Wheeler D.A., Krommiller B., Carlson J.W., Halpern A.,  
 RA Patel S., Adams M., Champagne M., Dugan S.P., Frise E., Hodson A.,  
 RA George R.A., Hoskins R.A., Laverly T., Muzny D.M., Nelson C.R.,  
 RA Pacle J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,  
 RA Svitskas R., Taber P.E., Wan K., Stapleton M., Sutton G., Venter C.,  
 RA Weinstein G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.,  
 RT "Finishing a whole-genome shotgun: release 3 of the *Drosophila*  
 RT *melanogaster* euchromatic genome sequence.";  
 RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).  
 [3]  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE=22426070; PubMed=12537573;  
 RA Kaminker J.S., Bergman C.M., Krommiller B., Carlson J., Svitskas R.,  
 RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,  
 RA Ashburner M., Celniker S.E.;  
 RT "The transposable elements of the *Drosophila melanogaster* euchromatin:  
 RT a genomic perspective.";  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 [4]  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE=22426069; PubMed=12537572;  
 RA Miya S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
 RA Hradecny P., Huang Y., Kaminker J.S., Millburn G.H., Prochuk S.E.,  
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,  
 RA Betencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,  
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,  
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
 RA Lewis S.E.;  
 RT "Annotation of the *Drosophila melanogaster* euchromatic genome: a  
 RT systematic review.";  
 RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).  
 [5]  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE=22426065; PubMed=12537568;  
 RA Celniker S.E., Wheeler D.A., Krommiller B., Carlson J.W., Halpern A.,  
 RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,  
 RA Ashburner M., Celniker S.E.;  
 RT "The transposable elements of the *Drosophila melanogaster* euchromatin:  
 RT a genomic perspective.";  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 [6]  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE=22426065; PubMed=12537568;  
 RA Celniker S.E., Wheeler D.A., Krommiller B., Carlson J.W., Halpern A.,  
 RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,  
 RA Ashburner M., Celniker S.E.;  
 RT "The transposable elements of the *Drosophila melanogaster* euchromatin:  
 RT a genomic perspective.";  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).

DR InterPro: IPR006209; EGF\_like.  
 DR Pfam: PF01826; TIL: 1.  
 DR PROSITE: PS01186; EGF\_2; UNKNOWN 1.  
 SQ SEQUENCE 119 AA; 13124 MW; 7A6021C71297D287 CRC64;  
 Query Match 88.9%; Score 32; DB 2; Length 119;  
 Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Oy 2 GCPV 6  
 Db 57 GCPV 61  
 RESULT 39  
 AAS65002 PRELIMINARY; PRT; 119 AA.  
 ID AAS65002;  
 AC AAS65002;  
 DT 01-APR-2004 (TrEMBLrel. 27, Created)  
 DT 01-APR-2004 (TrEMBLrel. 27, Last sequence update)  
 DT 01-APR-2004 (TrEMBLrel. 27, Last annotation update)  
 DE CG33259-PA.  
 GN CG33259.  
 OS *Drosophila melanogaster* (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; *Drosophila*.  
 OX NCBI\_TaxID=7227;  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amaratunga P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.H., Blazey R.G., Champagne M., Pfeiffer B.D.,  
 RA Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,  
 RA Abril J.F., Abmayyan A., An H.J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrera S., Fleischmann W.,  
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Gloder A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,  
 RA Jalaal M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusser D.R., Pacle J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svitskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Williams S.M., Woodger, Morley K.C., Wu D., Yang S., Yao Q.A., Ye J.,  
 RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195 (2000).  
 [2]  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE=22426065; PubMed=12537568;  
 RA Celniker S.E., Wheeler D.A., Krommiller B., Carlson J.W., Halpern A.,  
 RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,  
 RA Ashburner M., Celniker S.E.;  
 RT "The transposable elements of the *Drosophila melanogaster* euchromatin:  
 RT a genomic perspective.";  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).

RA Patel S., Adams M., Champe M., Dugan S.P., Friese E., Hodgson A.,  
 RA George R.A., Hoskins R.A., Lavery T., Muzny D.M., Nelson C.R.,  
 RA Paclel J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,  
 RA Svirskaas R., Tabor P.E., Wen K., Stapleton M., Sutton G.G., Venter C.,  
 RA Weinstein G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.,  
 RT "Finishing a whole-genome shotgun: release 3 of the Drosophila  
 RT melanogaster euchromatic genome sequence."  
 RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426069; PubMed=12537572;  
 RA Miura S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,  
 RA Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,  
 RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,  
 RA Bertencourt B.R., Celinker S.E., de Grey A.D., Drysdale R.A.,  
 RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,  
 RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
 RA Lewis S.E.,  
 RT "Annotation of the Drosophila melanogaster euchromatic genome: a  
 RT systematic review."  
 RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22426070; PubMed=12537573;  
 RA Kaminker J.S., Bergman C.M., Krommiller B., Carlson J., Svirskaas R.,  
 RA Patel S., Friese E., Wheeler D.A., Lewis S.E., Rubin G.M.,  
 RA Ashburner M., Celinker S.E.,  
 RT "The transposable elements of the Drosophila melanogaster euchromatin:  
 RT a genomic perspective."  
 RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RA FlyBase:  
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RA FlyBase:  
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AE003531; AAS65002.1; - 7A6021C71297D2B7 CRC64;  
 SQ SEQUENCE 119 AA; 13124 MW; 7A6021C71297D2B7 CRC64;  
 Query Match 88.9%; Score 32; DB 2; Length 119;  
 Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 GPCV 6  
 DB 57 GPCV 61

RESULT 40  
 Q6SHL1 PRELIMINARY; PRT; 132 AA.  
 AC Q6SHL1;  
 DT 05-JUL-2004 (TRENBLrel. 27, Created)  
 DT 05-JUL-2004 (TRENBLrel. 27, Last sequence update)  
 DT 05-JUL-2004 (TRENBLrel. 27, Last annotation update)  
 DE Hypothetical protein  
 GN ORFNames=EBAC750-09G06.25;  
 OS uncultured bacterium 314.  
 OC Bacteria; environmental samples.  
 CX NCB1\_TaxID=257387;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA DeLong E.F.,  
 RL Submitted (DEC-2003) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Heidelberg J.F., Eisen J.A., Nelson W.C., DeLong E.F.,  
 RL Submitted (NOV-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AY458634; AAR37610.1; -.  
 DR InterPro: IPR006683; Thioestr\_supf.  
 DR Pfam: PF03061; 4HBT; 1.

KW Hypothetical protein.  
 SQ SEQUENCE 132 AA; 14323 MW; 9359C91A9247C4E5 CRC64;  
 Query Match 88.9%; Score 32; DB 2; Length 132;  
 Best Local Similarity 100.0%; Pred. No. 3e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 GPCV 6  
 DB 71 GPCV 75

Search completed: December 30, 2004, 15:13:58  
 Job time : 8.93462 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: December 30, 2004, 14:59:21 ; Search time 0.211484 Seconds

(without alignments)  
2729.763 Million cell updates/sec

Title: US-10-719-385-26

Perfect score: 36

Sequence: 1 AGGPCV 6

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

1: PIR.79:.\*  
2: p1r1.\*  
3: p1r3.\*  
4: p1r4.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	36	100.0	407	2 T34535	hypothetical prote
2	36	100.0	444	2 F64420	hypothetical prote
3	36	100.0	469	2 C69460	conserved hypotet
4	35	97.2	383	2 S49928	F420-nonreducing h
5	35	97.2	383	2 S67477	F420-nonreducing h
6	35	97.2	501	2 H69119	hypothetical prote
7	35	97.2	1091	2 S33596	protein-tyrosine k
8	33	91.7	657	2 S32991	hypothetical prote
9	32	88.9	208	2 T41555	hypothetical prote
10	32	88.9	254	2 T12338	L-ascorbate peroxi
11	32	88.9	280	2 C70617	probable transcrip
12	32	88.9	292	2 S11870	peroxidase (EC 1.1
13	32	88.9	322	2 T03686	peroxidase (EC 1.1
14	32	88.9	374	2 JC2124	major allergen Cry
15	32	88.9	374	2 JC2124	major allergen Cry
16	32	88.9	386	2 A41950	retrovirus-related
17	32	88.9	398	2 C95278	hypothetical prote
18	32	88.9	404	2 T05556	pectate lyase (EC
19	32	88.9	404	2 S12209	pectate lyase (EC
20	32	88.9	407	2 S22586	homeotic protein E
21	32	88.9	426	2 F96994	uncharacterized co
22	32	88.9	437	2 JC4929	transcription fact
23	32	88.9	450	2 WABPT5	heme biosynthesis
24	32	88.9	454	2 A97181	gene D10 protein -
25	32	88.9	462	1 C71296	glycine-CRNA ligas
26	32	88.9	519	2 T43533	hexose transporter p
27	32	88.9	531	2 T40887	hexose transporter
28	32	88.9	535	2 T40887	hexose transporter
29	32	88.9	587	2 JC5669	Ca2+/calmodulin-de

30	32	88.9	687	2 T27421	hypothetical prote
31	32	88.9	699	2 C43674	US4 protein - huma
32	32	88.9	1220	2 A56136	tagged protein pre
33	32	88.9	1319	2 S55598	tegment protein 0
34	32	88.9	1408	2 S16148	gene serrate prote
35	32	88.9	2135	2 T14602	variant-specific s
36	31	86.1	48	2 S29216	neurotoxin Tx2 - s
37	31	86.1	49	2 S29216	neurotoxin Tx2 - s
38	31	86.1	291	2 T20083	hypothetical prote
39	31	86.1	291	2 T26576	hypothetical prote
40	31	86.1	570	4 B44282	retrovirus-related
41	31	86.1	2437	2 S42612	transmembrane prot
42	31	86.1	2824	2 T22759	hypothetical prote
43	30	83.3	96	2 S32182	hypothetical prote
44	30	83.3	104	2 T08518	hypothetical prote
45	30	83.3	104	2 J01757	hypothetical prote
46	30	83.3	213	2 S17815	bacteriophage phi
47	30	83.3	396	2 JC5022	UDP-galactose tran
48	29	80.6	44	2 S54144	tmal protein - mou
49	29	80.6	84	2 H96495	unknown protein, 4
50	29	80.6	101	2 A97623	hypothetical prote
51	29	80.6	105	2 A13464	hypothetical prote
52	29	80.6	110	2 G72540	hypothetical prote
53	29	80.6	111	2 I50622	Crot protein - chi
54	29	80.6	124	2 T48833	hypothetical prote
55	29	80.6	202	2 B45512	cold-regulated pro
56	29	80.6	208	2 T46927	hypothetical prote
57	29	80.6	209	2 C81691	conserved hypotet
58	29	80.6	210	2 S55671	hypothetical prote
59	29	80.6	244	2 F75520	indazoleglycerol-
60	29	80.6	255	2 C81301	hypothetical prote
61	29	80.6	273	2 T49602	related to homoti
62	29	80.6	346	2 I48185	gene alx3 protein
63	29	80.6	351	2 T31920	hypothetical prote
64	29	80.6	353	2 D69422	F420-nonreducing h
65	29	80.6	374	2 T22683	hypothetical prote
66	29	80.6	390	2 JC7162	Golgi UDP-galactos
67	29	80.6	432	2 A28913	regulatory protein
68	29	80.6	483	2 E49343	salicylaldehyde de
69	29	80.6	503	2 T48825	hypothetical prote
70	29	80.6	513	1 ACCHD1	nicotinic acetylch
71	29	80.6	554	2 C70512	hypothetical prote
72	29	80.6	662	2 D40228	neurexin II-beta p
73	29	80.6	673	4 F40201	artifac-warnin g
74	29	80.6	1715	2 C40228	neurexin II-alpha
75	29	80.6	2533	2 T28675	alpha-5ID immobili
76	29	80.6	2533	2 T28674	alpha-5ID immobili
77	29	80.6	3078	2 T28432	variant-specific s
78	29	80.6	4351	2 T00252	MEGFI protein - ra
79	28	77.8	52	2 G87562	hypothetical prote
80	28	77.8	63	2 S25772	testis-specific pr
81	28	77.8	64	2 C87540	hypothetical prote
82	28	77.8	68	2 S25775	testis-specific pr
83	28	77.8	74	2 S25773	testis-specific pr
84	28	77.8	75	2 AE1708	hypothetical prote
85	28	77.8	75	2 AH1337	hypothetical prote
86	28	77.8	80	1 S01498	pancreatic secreto
87	28	77.8	82	2 D81056	hypothetical prote
88	28	77.8	88	2 T47139	hypothetical prote
89	28	77.8	89	2 S71555	proteinase inhibit
90	28	77.8	89	2 J02361	wheat aluminium ind
91	28	77.8	90	2 A87553	Rebad protein (imp
92	28	77.8	103	2 C34770	ORF3 protein - sal
93	28	77.8	121	2 A10552	conserved hypotet
94	28	77.8	127	2 E87434	hypothetical prote
95	28	77.8	130	2 S41571	blaisticidin-S deam
96	28	77.8	134	2 T49623	hypothetical prote
97	28	77.8	138	2 S74831	hypothetical prote
98	28	77.8	141	2 G72641	hypothetical prote
99	28	77.8	143	2 B72627	hypothetical prote
100	28	77.8	154	2 H97860	hypothetical prote
101	28	77.8	171	2 B70408	conserved hypotet
102	28	77.8	172	1 RFMB	phycoerythrocyanin





C:/Date: 13-Jan-1995 #sequence\_revision 12-May-1995 #text\_change 09-Jul-2004  
C:/Accession: S49928; S67474  
R:/Depositor: U. Blaust, M.; Lentes, S.; Herzberg, C.; Gottschalk, G.  
A:/Description: Analysis of the vhcGAC and vhcGAC operons from Methanosarcina mazel strain  
A:/Reference number: S49928  
A:/Accession: S49928  
A:/Status: preliminary  
A:/Molecule type: DNA  
A:/Residues: 1-383 <DEP>  
A:/Cross-references: UNIPROT:Q50248; EMBL:X82940; NID:G599897; PIDN:CAA58113.1; PID:G5998  
R:/Depositor: U. Blaust, M.; Lentes, S.; Herzberg, C.; Gottschalk, G.  
Eur. J. Biochem. 227, 261-269, 1995  
A:/Title: Analysis of the vhcGAC and vhcGAC operons from Methanosarcina mazel strain Goel  
A:/Reference number: S67474; MUID:95154297; PMID:7851393  
A:/Accession: S67474  
A:/Status: preliminary  
A:/Molecule type: DNA  
A:/Residues: 1-383 <DEP>  
A:/Cross-references: EMBL:X82940; NID:G599897; PIDN:CAA58113.1; PID:G599898  
C:/Genetics:  
A:/Gene: vhcG  
C:/Superfamily: hydrogenase (NiFe) small chain

Query Match 97.2%; Score 35; DB 2; Length 383;  
Best Local Similarity 83.3%; Pred. No. 36;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGPCV 6  
Db 323 AGPCCI 328

RESULT 5  
S67477  
P420-nonreducing hydrogenase (EC 1.12.99.-) vhcG precursor - Methanosarcina mazel  
C:/Species: Methanosarcina mazel  
C:/Date: 14-Feb-1997 #sequence\_revision 13-Mar-1997 #text\_change 09-Jul-2004  
C:/Accession: S67477  
R:/Depositor: U. Blaust, M.; Lentes, S.; Herzberg, C.; Gottschalk, G.  
Eur. J. Biochem. 227, 261-269, 1995  
A:/Title: Analysis of the vhcGAC and vhcGAC operons from Methanosarcina mazel strain Goel  
A:/Reference number: S67474; MUID:95154297; PMID:7851393  
A:/Accession: S67477  
A:/Status: preliminary  
A:/Molecule type: DNA  
A:/Residues: 1-383 <DEP>  
A:/Cross-references: UNIPROT:Q50225; EMBL:X83112; NID:G602581; PIDN:CAA58176.1; PID:G6025  
C:/Superfamily: hydrogenase (NiFe) small chain  
C:/Keywords: oxidoreductase

Query Match 97.2%; Score 35; DB 2; Length 383;  
Best Local Similarity 83.3%; Pred. No. 36;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGPCV 6  
Db 323 AGPCCI 328

RESULT 6  
H69119  
Hypothetical protein MTH1895 - Methanobacterium thermoautotrophicum (strain Delta H)  
C:/Species: Methanobacterium thermoautotrophicum  
C:/Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 09-Jul-2004  
C:/Accession: H69119  
R:/Smith, D.R.; Doucette-Stamm, L.A.; Delonghery, C.; Lee, H.; Dubois, J.; Aldredge, T.;  
Kl., S.; Church, G.M.; Daniel, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.  
J. Bacteriol. 179, 7135-7155, 1997  
A:/Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H. func  
A:/Reference number: A69000; MUID:98037514; PMID:9371463  
A:/Accession: H69119

A:/Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:/Molecule type: DNA  
A:/Residues: 1-501 <MTH>  
A:/Cross-references: UNIPROT:Q27917; GB:AE000941; GB:AE000666; NID:G2623025; PIDN:AA8635;  
A:/Experimental source: strain Delta H  
C:/Genetics:  
A:/Gene: MTH1895  
A:/Start codon: TTG  
C:/Superfamily: hypothetical protein M0966

Query Match 97.2%; Score 35; DB 2; Length 501;  
Best Local Similarity 83.3%; Pred. No. 46;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGPCV 6  
Db 94 AGPCCI 99

RESULT 7  
S33596  
Protein-tyrosine kinase (EC 2.7.1.112) - human  
C:/Species: Homo sapiens (man)  
C:/Date: 02-Dec-1993 #sequence\_revision 01-Sep-1995 #text\_change 16-Aug-2004  
C:/Accession: S33596  
R:/Manser, E.; Leung, T.; Salihuddin, H.; Tan, L.; Lim, L.  
Nature 363, 364-367, 1993  
A:/Title: A non-receptor tyrosine kinase that inhibits the GTPase activity of p21 (cdc42).  
A:/Reference number: S33596; MUID:93268389; PMID:8497321  
A:/Accession: S33596  
A:/Status: preliminary  
A:/Molecule type: mRNA  
A:/Residues: 1-1091 <MAN>  
A:/Cross-references: UNIPROT:Q07912; EMBL:L13738; NID:G307304; PIDN:AAA53570.1; PID:G30730  
C:/Superfamily: protein kinase homology  
C:/Keywords: ATP; phosphotransferase; tyrosine-specific protein kinase  
F188-196/Region: protein kinase ATP-binding motif

Query Match 97.2%; Score 35; DB 2; Length 1091;  
Best Local Similarity 83.3%; Pred. No. 89;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGPCV 6  
Db 892 AGPCCI 897

RESULT 8  
S32991  
Hypothetical protein - human herpesvirus 4  
C:/Species: human herpesvirus 4, Epstein-Barr virus  
C:/Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 13-Jan-1995  
C:/Accession: S32991  
R:/Farrell, P.J.  
submitted to the EMBL Data Library, March 1988  
A:/Reference number: S32973  
A:/Accession: S32991  
A:/Status: preliminary  
A:/Molecule type: DNA  
A:/Residues: 1-657 <PAR>  
A:/Cross-references: EMBL:V01555

Query Match 91.7%; Score 33; DB 2; Length 657;  
Best Local Similarity 83.3%; Pred. No. 1,4e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGPCV 6  
Db 399 AGPCCI 404

RESULT 9

T41555  
 hypothetical protein SPCC70.09c - fission yeast (Schizosaccharomyces pombe)  
 C/Species: Schizosaccharomyces pombe  
 C/Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
 C/Accession: T41555  
 R/Mod: V.; Rajandream, M.A.; Barrell, B.G.; Seeger, K.; Harris, D.  
 Submitted to the EMBL Data Library, June 1998  
 A/Reference number: 222001  
 A/Accession: T41555  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: DNA  
 A/Residues: 1-208 <MO>  
 A/Cross-references: UNIPROT:O94243; EMBL:AL023794; PIDN:CAA13359.1; GSPDB:GN00068; SPDB:  
 A/Experimental source: strain 972h-; cosmid c70  
 A/Genetics:  
 A/Gene: SPDB:SPCC70.09c  
 A/Map position: 3  
 A/Introns: 48/1

Query Match  
 Best Local Similarity 86.9%; Score 32; DB 2; Length 208;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGPC 5  
 |||||  
 DB 5 AGGPC 9

## RESULT 10

T12338  
 L-ascorbate peroxidase (EC 1.11.1.11) - common ice plant  
 C/Species: Mesembryanthemum crystallinum (common ice plant)  
 C/Date: 23-Jul-1998 #sequence\_revision 23-Jul-1998 #text\_change 12-Jul-2004  
 C/Accession: T12338  
 R/Michalowski, C.B.; Bohner, H.J.  
 Submitted to the EMBL Data Library, July 1998  
 A/Description: A putative ascorbate peroxidase from the common ice plant Mesembryanthemum  
 A/Reference number: Z17496  
 A/Accession: T12338  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: mRNA  
 A/Residues: 1-254 <MIC>  
 A/Cross-references: UNIPROT:O81604; EMBL:AF079513; NID:G3377754; PID:G3377755  
 C/Superfamily: peroxidase  
 C/Keywords: heme; iron; metalloprotein; oxidoreductase  
 F/160/Binding site: heme iron (His) (proximal axial ligand) #status predicted

Query Match  
 Best Local Similarity 88.9%; Score 32; DB 2; Length 254;  
 Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGPCV 6  
 |||||  
 DB 107 SGGPCI 112

## RESULT 11

C70617  
 probable transcription regulator RV0144 - Mycobacterium tuberculosis (strain H37RV)  
 C/Species: Mycobacterium tuberculosis  
 C/Date: 17-Jul-1998 #sequence\_revision 17-Jul-1998 #text\_change 09-Jul-2004  
 C/Accession: C70617  
 R/Cole, S.T.; Broach, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltham, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.  
 Nature 393, 537-544, 1998  
 A/Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.  
 A/Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome  
 A/Reference number: A70500; PMID:98295987; PMID:9634230  
 A/Accession: C70617  
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A/Molecule type: DNA  
 A/Residues: 1-280 <COL>

A/Cross-references: UNIPROT:P96821; GB:Z92770; GB:AL123456; NID:G3261720; PIDN:CAB07050.1  
 A/Experimental source: strain H37RV  
 A/Genetics:  
 A/Gene: RV0144

Query Match  
 Best Local Similarity 88.9%; Score 32; DB 2; Length 280;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGPC 5  
 |||||  
 DB 33 AGGPC 37

## RESULT 12

S11870  
 peroxidase (EC 1.11.1.7) - cucumber (fragment)  
 C/Species: Cucumis sativus (cucumber)  
 C/Date: 30-Jun-1992 #sequence\_revision 30-Jun-1992 #text\_change 09-Jul-2004  
 C/Accession: S11870  
 R/Morgans, P.H.; Callahan, A.M.; Dunn, L.J.; Abeles, F.B.  
 Plant Mol. Biol. 14, 715-725, 1990  
 A/Title: Isolation and sequencing of cDNA clones encoding ethylene-induced putative pero  
 A/Reference number: S11870; PMID:9134662; PMID:2102850  
 A/Accession: S11870  
 A/Molecule type: mRNA  
 A/Residues: 1-292 <MOR>

A/Cross-references: UNIPROT:P19135; GB:M32742; NID:G167516; PIDN:AAA3121.1; PID:G167517  
 A/Note: 231-Leu was also found  
 C/Superfamily: peroxidase  
 C/Keywords: chromoprotein; glycoprotein; heme; iron; metalloprotein; oxidoreductase  
 F/7-86/Disulfide bonds: #status predicted  
 F/34/Active site: Arg #status predicted  
 F/38,164/Binding site: heme iron (His) (axial ligands) #status predicted  
 F/40-45/Disulfide bonds: #status predicted  
 F/66,119,179/Binding site: carbonylate (Asn) (covalent) #status predicted  
 F/92-288/Disulfide bonds: #status predicted  
 F/171-199/Disulfide bonds: #status predicted

Query Match  
 Best Local Similarity 88.9%; Score 32; DB 2; Length 292;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGPC 5  
 |||||  
 DB 107 AGGPC 111

## RESULT 13

T03686  
 peroxidase (EC 1.11.1.7) - common tobacco  
 C/Species: Nicotiana tabacum (common tobacco)  
 C/Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 09-Jul-2004  
 C/Accession: T03686  
 R/Osakabe, K.; Kawai, S.; Katayama, Y.; Morohoshi, N.  
 Submitted to the EMBL Data Library, June 1992  
 A/Description: Nucleotide sequence for the genomic DNA encoding the anionic peroxidase ge  
 A/Reference number: Z15008  
 A/Accession: T03686  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: DNA  
 A/Residues: 1-322 <OSA>  
 A/Cross-references: UNIPROT:Q42964; EMBL:D11396; PIDN:BA01992.1  
 A/Experimental source: strain NK326  
 A/Genetics:

A/Introns: 66/3; 128/3; 184/1  
 C/Superfamily: peroxidase  
 C/Keywords: chromoprotein; heme; iron; metalloprotein; oxidoreductase  
 F/33-109/Disulfide bonds: #status predicted  
 F/60/Active site: Arg #status predicted  
 F/64,187/Binding site: heme iron (His) (axial ligands) #status predicted  
 F/66-69/Disulfide bonds: #status predicted  
 F/115-318/Disulfide bonds: #status predicted

F;194-226/D1enulfide bonds: #status predicted

Query Match 88.9%; Score 32; DB 2; Length 322;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGGPC 5  
|||||  
Db 130 AGGPC 134

RESULT 14

JC2123  
major allergen Cry j I precursor (clone PCCI-2-2) - Japanese cedar

C/Species: Cryptomeria japonica (Japanese cedar)  
C/Date: 14-Jul-1994 #sequence\_revision 14-Jul-1994 #text\_change 09-Jul-2004  
C/Accession: JC2123; PC2065  
R/Sone, T.; Komiyama, N.; Shimizu, K.; Kusakabe, T.; Morikubo, K.; Kino, K.  
Biochem. Biophys. Res. Commun. 199, 619-625, 1994  
A/Title: Cloning and sequencing of cDNA coding for Cry j I, a major allergen of Japanese  
A/Reference number: JC2123; MUID:94183234; PMID:8135802  
A/Accession: JC2123  
A/Molecule type: mRNA  
A/Residues: 1-374 <SON>  
A/Cross-references: UNIPROT:P18632; GB:D26544; NID:G493631; PID:BA05542.1; PID:G493632  
A/Experimental source: pollen

A/Accession: PC2065  
A/Molecule type: protein  
A/Residues: 22-53;58-81;219-232;236-258;299-307;346-372 <SO2>  
A/Note: the authors described carbohydrate binding site for residue 279  
C/Superfamily: pectate lyase LAT59  
C/Keywords: glycoprotein; pollen  
F;1-21/Domain: signal sequence #status predicted <SIG>  
F;22-374/Product: major allergen Cry j I (clone PCCI-2-2) #status predicted <MAT>  
F;158,191,293,354/Binding site: carbohydrate (Aan) (covalent) #status predicted

Query Match 88.9%; Score 32; DB 2; Length 374;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GGPCV 6  
|||||  
Db 125 GGPCV 129

RESULT 15

JC2124  
major allergen Cry j I precursor (clone PCCI-15) - Japanese cedar

C/Species: Cryptomeria japonica (Japanese cedar)  
C/Date: 14-Jul-1994 #sequence\_revision 14-Jul-1994 #text\_change 09-Jul-2004  
C/Accession: JC2124  
R/Sone, T.; Komiyama, N.; Shimizu, K.; Kusakabe, T.; Morikubo, K.; Kino, K.  
Biochem. Biophys. Res. Commun. 199, 619-625, 1994  
A/Title: Cloning and sequencing of cDNA coding for Cry j I, a major allergen of Japanese  
A/Reference number: JC2123; MUID:94183234; PMID:8135802  
A/Accession: JC2124  
A/Molecule type: mRNA  
A/Residues: 1-374 <SON>  
A/Cross-references: UNIPROT:P18632; GB:D26544; NID:G493633; PID:BA05543.1; PID:G493634  
A/Experimental source: pollen  
A/Note: the authors described carbohydrate binding site for residue 279  
C/Superfamily: pectate lyase LAT59  
C/Keywords: glycoprotein; pollen  
F;1-21/Domain: signal sequence #status predicted <SIG>  
F;22-374/Product: major allergen Cry j I (clone PCCI-15) #status predicted <MAT>  
F;158,191,293,354/Binding site: carbohydrate (Aan) (covalent) #status predicted

Query Match 88.9%; Score 32; DB 2; Length 374;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GGPCV 6  
|||||

Db 125 GGPCV 129

RESULT 16

A41950  
retrovirus-related hypothetical protein 1 - Trypanosoma cruzi retrotransposon

C/Species: Trypanosoma cruzi  
C/Date: 04-Mar-1993 #sequence\_revision 18-Nov-1994 #text\_change 08-Jan-1999  
C/Accession: A41950  
R/Villanueva, M.S.; Williams, S.P.; Beard, C.B.; Richards, F.F.; Aksoy, S.  
Mol. Cell. Biol. 11, 6139-6148, 1991  
A/Title: A new member of a family of site-specific retrotransposons is present in the spj  
A/Reference number: A41950; MUID:92049344; PMID:1179380  
A/Accession: A41950  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-386 <VIL>  
A/Cross-references: GB:M62862; NID:G162247; PID:G162248  
A/Note: sequence extracted from NCBI backbone (NCBI:66378, NCBI:66379)

Query Match 88.9%; Score 32; DB 2; Length 386;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGGPC 5  
|||||  
Db 346 AGGPC 350

RESULT 17

C95278

hypothetical protein Sma0247 [imported] - Sinorhizobium meliloti (strain 1021) magaplaasmu

C/Species: Sinorhizobium meliloti  
C/Date: 24-Aug-2001 #sequence\_revision 24-Aug-2001 #text\_change 09-Jul-2004  
C/Accession: C95278  
R/Barnett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; Bowser,  
J.; Kalman, S.; Keating, D.H.; Palm, S.R.; Peck, M.C.; Surzycki, R.; Welle, D.H.; Yeh, K.C.;  
Proc. Natl. Acad. Sci. U.S.A. 98, 9883-9888, 2001  
A/Title: Nucleotide sequence and predicted functions of the entire Sinorhizobium meliloti  
A/Reference number: A95262; MUID:21396509; PMID:11481432  
A/Accession: C95278  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-398 <KUR>  
A/Cross-references: UNIPROT:Q930R4; GB:AE006469; PID:AAK64789.1; PID:G14523198; GSPDB:GT  
A/Experimental source: strain 1021, megaplaasmid pSYMA  
R/Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,  
pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;  
L.; Hyman, R.W.; Jones, T.  
Science 293, 668-672, 2001

A/Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,  
heault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yen, K.C.  
A/Title: The composite genome of the legume symbiont Sinorhizobium meliloti.  
A/Reference number: A96039; MUID:21368234; PMID:11474104  
A/Contents: annotation  
C/Genetics:  
A/Gene: Sma0247  
A/Genome: plasmid

Query Match 88.9%; Score 32; DB 2; Length 398;  
Best Local Similarity 100.0%; Pred. No. 1.3e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGGPC 5  
|||||  
Db 37 AGGPC 41

RESULT 18

T05556  
pectate lyase (EC 4.2.2.2) F22X18.20 - Arabidopsis thaliana  
N/Alternate names: protein F22X18.20  
C/Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 23-Apr-1999 #sequence\_revision 23-Apr-1999 #text\_change 23-Jul-1999  
C/Accession: T05556  
R:Bevan, M.; Wedler, H.; Wedler, E.; Wambutt, R.; Hohnsiegel, J.; Mewes, H.W.; Mayer, K.F.  
Submitted to the Protein Sequence Database, February 1999  
A/Reference number: T05556  
A/Accession: T05556  
A/Molecule type: DNA  
A/Residues: 1-404 <BEV>  
A/Cross-references: EMBL:AL035356  
A/Experimental source: cultivar Columbia; BAC clone F22K18  
C/Genetics:  
A/Map position: 4  
A/Introns: 37/2; 287/2  
A/Note: F22K18.20  
C/Superfamily: pectate lyase LATS9  
C/Keywords: carbon-oxygen lyase

Query Match 88.9%; Score 32; DB 2; Length 404;  
Best Local Similarity 66.7%; Pred. No. 1.4e+02;  
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGPCV 6  
Db 159 SGGPCI 164

RESULT 19  
S12209  
pectate lyase (EC 4.2.2.2) - tomato  
C/Species: Lycopersicon esculentum (tomato)  
C/Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 09-Jul-2004  
C/Accession: S12209  
R:Budeller, K.A.; Smith, A.G.; Gasser, C.S.  
Mol. Gen. Genet. 224, 183-192, 1990  
A/Title: Regulation of a stylar transmitting tissue-specific gene in wild-type and trans  
A/Reference number: S12209; MUID:91117185; PMID:2277637  
A/Accession: S12209  
A/Status: preliminary  
A/Molecule type: mRNA  
A/Residues: 1-404 <BD>  
A/Cross-references: UNIPROT:P24396; GB:X55193; NID:919161; PIDN:CA38979.1; PID:919162  
C/Superfamily: pectate lyase LATS9  
C/Keywords: carbon-oxygen lyase

Query Match 88.9%; Score 32; DB 2; Length 404;  
Best Local Similarity 66.7%; Pred. No. 1.4e+02;  
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGPCV 6  
Db 155 SGGPCI 160

RESULT 20  
S22586  
homeotic protein EVX1 - human  
C/Species: Homo sapiens (man)  
C/Date: 19-Feb-1994 #sequence\_revision 10-Nov-1995 #text\_change 16-Aug-2004  
C/Accession: S22586  
R:Faella, A.; d'Esposito, M.; Rambaldi, M.; Acampora, D.; Balsaniore, S.; Stornaiuolo,  
Nucleic Acids Res. 19, 6541-6545, 1991  
A/Title: Isolation and mapping of EVX1, a human homeobox gene homologous to even-skipped  
A/Reference number: S22586; MUID:92093615; PMID:1684419  
A/Accession: S22586  
A/Status: preliminary  
A/Molecule type: mRNA  
A/Residues: 1-407 <FA>  
A/Cross-references: UNIPROT:P49640; EMBL:X60655; NID:931278; PIDN:CAA43062.1; PID:977357  
C/Superfamily: homeobox homology  
C/Keywords: DNA binding; homeobox; nucleus; transcription regulation  
F:184-240/Domain: homeobox homology <HOX>

Query Match 88.9%; Score 32; DB 2; Length 407;

Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGPCV 5  
Db 343 AGGPC 347

RESULT 21  
F96994  
uncharacterized conserved protein CAC0769 [imported] - Clostridium acetobutylicum  
C/Species: Clostridium acetobutylicum  
C/Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 09-Jul-2004  
C/Accession: F96994  
R:Nojima, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,  
Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
J. Bacteriol. 183, 4823-4838, 2001  
A/Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clo  
A/Reference number: A96900; MUID:21359325; PMID:21359325  
A/Accession: F96994  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-746 <KUR>  
A/Cross-references: UNIPROT:Q97KX7; GB:AE001437; PIDN:AAK78745.1; PID:915023654; GSPDB:G  
A/Experimental source: Clostridium acetobutylicum ATCC824  
C/Genetics:  
A/Gene: CAC0769

Query Match 88.9%; Score 32; DB 2; Length 426;  
Best Local Similarity 66.7%; Pred. No. 1.4e+02;  
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGPCV 6  
Db 151 SGGPCI 156

RESULT 22  
JC4929  
transcription factor E2F1 - human  
N/Alternate names: Ap12 protein; retinoblastoma-associated protein; retinoblastoma-bindin  
C/Species: Homo sapiens (man)  
C/Date: 22-Oct-1996 #sequence\_revision 01-Nov-1996 #text\_change 09-Jul-2004  
C/Accession: JC4929; A45032; A42998; A42997; I54091  
R:Neuman, E.; Sellers, W.R.; McNeil, J.A.; Lawrence, J.B.; Kaelin Jr., W.G.  
Gene 173, 163-169, 1996  
A/Title: Structure and partial genomic sequence of the human E2F1 gene.  
A/Reference number: JC4929; MUID:97082961; PMID:8964493  
A/Accession: JC4929  
A/Status: preliminary; nucleic acid sequence not shown  
A/Molecule type: DNA  
A/Residues: 1-437 <NEU>  
A/Cross-references: UNIPROT:Q01094; GB:U47675; NID:91594281; GB:U47676; NID:91594282; GB  
A/Experimental source: placenta  
R:Shan, B.; Zhu, X.; Chen, P.L.; Durfee, T.; Yang, Y.; Sharp, D.; Lee, W.H.  
Mol. Cell. Biol. 12, 5620-5631, 1992  
A/Title: Molecular cloning of cellular genes encoding retinoblastoma-associated proteins  
A/Reference number: A45032; MUID:93078763; PMID:1448092  
A/Accession: A45032  
A/Molecule type: mRNA  
A/Residues: 1-CRQRRPAERDRLAALACRGAPPPPTARRANVSV, 1-312, 'S', 314-321, 'N', 323-328, 'T',  
A/Cross-references: GB:549592; NID:9260573; PIDN:AA824289.1; PID:9260574  
A/Note: sequence extracted from NCBI backbone (NCBI:P119095)  
A/Note: the authors are uncertain whether Met-1 is the initiator or whether translation;  
R:Helin, K.; Lees, J.A.; Vidal, M.; Dyson, N.; Harlow, E.; Fattaei, A.  
Cell 70, 337-350, 1992  
A/Title: A cDNA encoding a PRB-binding protein with properties of the transcription fact  
A/Reference number: A42998; MUID:92346720; PMID:1638654  
A/Accession: A42998  
A/Molecule type: mRNA  
A/Residues: 1-312, 'S', 314-321, 'N', 323-328, 'T', 330-437 <HEL>  
A/Cross-references: GB:M96577; NID:9181917; PIDN:AA35782.1; PID:9181918  
A/Experimental source: Nalm 6 pre-B leukemia cell line

A/Note: sequence extracted from NCBI backbone (NCBIN:110015, NCBI:110016)  
 R/Kellin Jr., W.G.; Krek, W.; Sellers, W.R.; Decaprio, J.A.; Ajchenbaum, F.; Fuchs, C.S.;  
 Cell 70, 351-364, 1992  
 A/Title: Expression cloning of a cDNA encoding a retinoblastoma-binding protein with E2F  
 A/Reference number: A42997; MUID:92346721; PMID:1638635  
 A/Accession: A42997  
 A/Molecule type: mRNA  
 A/Residues: 1-312, 'S', 314-321, 'N', 323-328, 'T', 330-437 <KAB>  
 A/Cross-references: GB:U13851; NID:9585713  
 A/Experimental source: Akata cells; expression vector pGEX-2TK  
 A/Note: sequence extracted from NCBI backbone (NCBIN:110018, NCBI:110019)  
 R/Johnson, D.G.; Ohnishi, K.; Nevins, J.R.  
 Genes Dev. 8, 1514-1525, 1994  
 A/Title: Autoregulatory control of E2F1 expression in response to positive and negative  
 A/Reference number: 154091; MUID:95047311; PMID:7958836  
 A/Accession: 154091  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: DNA  
 A/Residues: 1-88, 'R', 'T', 122-123, 'TPR', 127, 'ORR', 297-299, 'PRR', 308-309, 'RA', 312, 'C' <RES>  
 A/Cross-references: GB:674230; NID:9712816; PIDN:AD14150.1; PID:94261850  
 C/Genetics:  
 A/Gene: GDB:E2F1  
 A/Cross-references: GDB:134661; OMIM:189971  
 A/Map position: 20q11-20q11  
 A/Introns: 87/3 118/1 191/2 242/2 280/3 356/1  
 C/Keywords: DNA binding; transcription factor  
 F/67-108/Region: cyclin box #status predicted  
 F/118-190/Domain: DNA binding #status predicted <DNA>  
 F/191-241/Region: 7-residue repeats

Query Match 88.9%; Score 32; DB 2; Length 437;  
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGPC 5  
 |||||  
 Db 8 AGPC 12

RESULT 23  
 WABPT5  
 gene D10 protein - phage T5  
 C/Species: phage T5  
 A/Note: host Escherichia coli  
 C/Date: 30-Sep-1991 #sequence\_revision 30-Sep-1991 #text\_change 09-Jul-2004  
 C/Accession: S01931  
 R/Kaliman, A.V.; Kryukov, V.M.; Bayev, A.A.  
 Nucleic Acids Res. 16, 10353-10354, 1988  
 A/Title: The nucleotide sequence of the region of bacteriophage T5 early genes D10-D15.  
 A/Reference number: S01931; MUID:89057468; PMID:3057441  
 A/Accession: S01931  
 A/Status: translation not shown  
 A/Molecule type: DNA  
 A/Residues: 1-450 <KAL>  
 A/Cross-references: UNIPROT:P11107; EMBL:X12930; NID:915407; PIDN:CAA31398.1; PID:915405  
 C/Genetics:  
 A/Gene: D10  
 C/Superfamily: phage T5 gene D10 protein  
 C/Keywords: DNA binding; early protein

Query Match 88.9%; Score 32; DB 1; Length 450;  
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GPCV 6  
 |||||  
 Db 347 GPCV 351

RESULT 24  
 A97181  
 heme biosynthesis (nitro-2) family protein [imported] - Clostridium acetobutylicum  
 C/Species: Clostridium acetobutylicum

C/Date: 14-Sep-2001 #sequence\_revision 14-Sep-2001 #text\_change 09-Jul-2004  
 C/Accession: A97181  
 R/Nolling, J.; Brelton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,  
 J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.  
 J. Bacteriol. 183, 4823-4838, 2001  
 A/Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clostridium acetobutylicum  
 A/Reference number: A96900; MUID:21359325; PMID:21359325  
 A/Accession: A97181  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-454 <KUR>  
 A/Cross-references: UNIPROT:Q97GTE; GB:AE001437; PIDN:AAK80236.1; PID:915025284; GSPDB:GN  
 A/Experimental source: Clostridium acetobutylicum ATCC824  
 C/Genetics:  
 A/Gene: CAC2279

Query Match 88.9%; Score 32; DB 2; Length 454;  
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GPCV 6  
 |||||  
 Db 335 GPCV 339

RESULT 25  
 C71296  
 glycine-tRNA ligase (EC 6.1.1.14) (glys) - syphilis spirochete  
 C/Species: Treponema pallidum subsp. pallidum (syphilis spirochete)  
 C/Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 09-Jul-2004  
 C/Accession: C71296  
 R/Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwin  
 rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Uterback, T.; McDor  
 they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.  
 Science 281, 375-388, 1998  
 A/Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.  
 A/Reference number: A71250; MUID:98332770; PMID:9665876  
 A/Accession: C71296  
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A/Molecule type: DNA  
 A/Residues: 1-462 <COL>  
 A/Cross-references: UNIPROT:O83678; GB:AE001241; GB:AE000520; NID:93322969; PIDN:AAK26570  
 A/Experimental source: strain Nichols  
 C/Genetics:  
 A/Gene: TP0672  
 C/Superfamily: Mycoplasma genitalium glycine-tRNA ligase  
 C/Keywords: aminoacyl-tRNA synthetase; ligase; protein biosynthesis

Query Match 88.9%; Score 32; DB 1; Length 462;  
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGPC 5  
 |||||  
 Db 103 AGPC 107

RESULT 26  
 T43533  
 hexose transport protein Glt2 - fission yeast (Schizosaccharomyces pombe)  
 C/Species: Schizosaccharomyces pombe  
 C/Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 09-Jul-2004  
 C/Accession: T43533  
 R/Helland, S.; Hofer, M.; Radovanovic, N.; Lichtenberg-Frate, H.  
 submitted to the EMBL Data Library, August 1997  
 A/Description: Multiple hexose transporters of Schizosaccharomyces pombe.  
 A/Reference number: 222549  
 A/Accession: T43533  
 A/Status: translated from GB/EMBL/DBJ  
 A/Molecule type: mRNA  
 A/Residues: 1-519 <HEI>  
 A/Cross-references: UNIPROT:O74969; EMBL:AF017180; PIDN:AAB70519.1  
 A/Experimental source: strain 972h(-)

C/Genetics:  
A/Gene: ghtc2  
C/Superfamily: maltose transport protein MAL61

Query Match 88.9%; Score 32; DB 2; Length 519;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGPC 5  
|||||  
Db 240 AGGPC 244

## RESULT 27

T40480  
hexose transporter - fission yeast (Schizosaccharomyces pombe)  
C/Species: Schizosaccharomyces pombe  
C/Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
C/Accession: T40480  
R/Beck, A.; Reinhardt, R.; Lyne, M.; Wood, V.; Rajandream, M.A.; Barrell, B.G.  
submitted to the EMBL Data Library, May 1997  
A/Reference number: Z21932  
A/Accession: T40480  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-531 <BEC>  
A/Cross-references: UNIPROT:O74969; EMBL:AL023706; PIDN:CAA19288.1; GSPDB:GN00067; SPDB:  
C/Genetics:  
A/Gene: ghtc2; SPDB:SPBC44.08  
A/Map position: 2  
C/Superfamily: maltose transport protein MAL61

Query Match 88.9%; Score 32; DB 2; Length 511;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGPC 5  
|||||  
Db 252 AGGPC 256

## RESULT 28

T40887  
hexose transporter ghtc6 [similarity] - fission yeast (Schizosaccharomyces pombe)  
C/Species: Schizosaccharomyces pombe  
C/Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
C/Accession: T40887; T50484  
R/Mood, V.; Rajandream, M.A.; Barrell, B.G.; Murphy, L.; Harris, D.  
submitted to the EMBL Data Library, September 1998  
A/Reference number: Z21954  
A/Accession: T40887  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-535 <MOO>

A/Cross-references: UNIPROT:O74849; EMBL:AL031764; PIDN:CAA21117.1; GSPDB:GN00068; SPDB:  
A/Experimental source: strain 972h(-); cosmid c1235  
R/Helland, S.; Hoefler, M.; Radovanovic, N.; Winderickx, J.; Lichtenberg, H.  
submitted to the EMBL Data Library, October 1998  
A/Description: Multiple hexose transporters of Schizosaccharomyces pombe.  
A/Reference number: Z25086  
A/Accession: T50484  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-535 <HEI>  
A/Cross-references: EMBL:AF098076; PIDN:AAC64976.1  
A/Experimental source: strain 972h(-)  
C/Genetics:  
A/Gene: SPDB:SPCC1235.13; ghtc  
A/Map position: 3  
C/Superfamily: maltose transport protein MAL61

Query Match 88.9%; Score 32; DB 2; Length 535;

Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGPC 5  
|||||  
Db 250 AGGPC 254

## RESULT 29

JC5669  
Ca2+/calmodulin-dependent protein kinase kinase (EC 2.7.1.-) beta chain - rat  
C/Species: Rattus norvegicus (Norway rat)  
C/Date: 20-Nov-1997 #sequence\_revision 20-Nov-1997 #text\_change 16-Aug-2004  
C/Accession: JC5669; PC4493  
R/Kitani, T.; Okuno, S.; Fujisawa, H.  
J. Biochem. 122, 243-250, 1997  
A/Title: Molecular cloning of Ca2+/calmodulin-dependent protein kinase kinase beta.  
A/Reference number: JC5669; MIMD:97420710; PMID:9276695  
A/Accession: JC5669  
A/Molecule type: mRNA  
A/Residues: 1-587 <KIT>  
A/Cross-references: UNIPROT:O88831; GB:AB018081; NID:G3702720; PIDN:BAA33524.1; PID:d103;  
A/Experimental source: cerebellar  
A/Accession: PC4493  
A/Molecule type: protein  
A/Residues: 425-501 <KIT>

A/Experimental source: cerebellar  
C/Superfamily: protein kinase homology  
C/Keywords: ATP; phosphotransferase  
F/162-445/Domain: protein kinase homology <KIN>  
F/170-194/Domain: ATP-binding #status predicted <ATP>  
F/480-493/Domain: calmodulin-binding #status predicted <CAB>

Query Match 88.9%; Score 32; DB 2; Length 587;  
Best Local Similarity 100.0%; Pred. No. 1.9e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGPCV 6  
|||||  
Db 557 GGPCV 561

## RESULT 30

T27421  
hypothetical protein Y76A2B.3 - Caenorhabditis elegans  
C/Species: Caenorhabditis elegans  
C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C/Accession: T27421  
R/Steward, C.  
submitted to the EMBL Data Library, October 1998  
A/Reference number: Z20363  
A/Accession: T27421  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-667 <WIL>

A/Cross-references: UNIPROT:Q9XMD1; EMBL:AL032658; PIDN:CAA21744.1; GSPDB:GN00021; CESP:  
A/Experimental source: clone Y76A2B  
C/Genetics:  
A/Gene: CESP:Y76A2B.3  
A/Map position: 3  
A/Intons: 46/3; 84/3; 130/1; 220/3; 363/3; 494/3; 582/2  
C/Superfamily: human long-chain-fatty-acid-CoA ligase; acetate-CoA ligase homology  
F/126-669/Domain: acetate-CoA ligase homology <ACL>

Query Match 88.9%; Score 32; DB 2; Length 687;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGPCV 6  
|||||  
Db 88 GGPCV 92

## RESULT 31

C43674  
USA protein - human herpesvirus 2 (strain HG52)  
C:Species: human herpesvirus 2  
C:Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 09-Jul-2004  
C:Accession: C43674  
R:McGeoch, D.J.; Moss, H.W.M.; McNab, D.; Frame, M.C.  
J. Gen. Virol. 68, 19-38, 1987  
A:Title: DNA sequence and genetic content of the HindIII 1 region in the short unique co  
nary comparisons.  
A:Reference number: A43674; MUID:87111457; PMID:3027242  
A:Accession: C43674  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-659 <MG>  
A:Cross-references: UNIPROT:P13290; EMBL:X04798; NID:G53900; PID:CAA28490.1; PID:G53904  
C:Superfamily: collagen alpha 1(VIII) chain; complement C1q carboxyl-terminal homology

Query Match 88.9%; Score 32; DB 2; Length 699;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 GPCV 6  
|||||

Db 215 GPCV 219

## RESULT 32

A56136  
Jagged protein precursor - rat  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 28-Apr-1995 #sequence\_revision 28-Apr-1995 #text\_change 21-Jul-2003  
C:Accession: A56136  
R:Lindseil, C.E.; Shawber, C.J.; Boulter, J.; Weinmaster, G.  
Cell 80, 909-917, 1995  
A:Title: Jagged: a mammalian ligand that activates Notch1.  
A:Reference number: A56136; MUID:95211842; PMID:7697721  
A:Accession: A56136  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-1220 <LIN>  
A:Cross-references: GB:L38483  
F:379-410/Domain: EGF homology <EGF1>  
F:492-523/Domain: EGF homology <EGF>  
F:634-665/Domain: EGF homology <EGF2>

Query Match 88.9%; Score 32; DB 2; Length 1220;  
Best Local Similarity 100.0%; Pred. No. 3.5e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGPC 5  
|||||

Db 89 AGPC 93

## RESULT 33

S55598  
Legument protein 03 - equine herpesvirus 2  
C:Species: equine herpesvirus 2  
C:Date: 27-Oct-1995 #sequence\_revision 03-Nov-1995 #text\_change 09-Jul-2004  
C:Accession: S55598  
R:Telford, E.A.R.; Watson, M.S.; Aird, H.C.; Perry, J.; Davison, A.J.  
J. Mol. Biol. 249, 520-528, 1995  
A:Title: The DNA sequence of equine herpesvirus 2.  
A:Reference number: S55594; MUID:95302501; PMID:7783207  
A:Accession: S55598  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-1319 <TEL>  
A:Cross-references: UNIPROT:Q66609; GB:U20824; NID:G695172; PID:AA313791.1; PID:G695176  
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, February 1995

Query Match 88.9%; Score 32; DB 2; Length 1319;

Best Local Similarity 100.0%; Pred. No. 3.8e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGPC 5  
|||||

Db 116 AGPC 120

## RESULT 34

S16148  
gene serrate protein precursor - fruit fly (Drosophila melanogaster)  
C:Species: Drosophila melanogaster  
C:Date: 31-Dec-1991 #sequence\_revision 02-Aug-1994 #text\_change 09-Jul-2004  
C:Accession: S16148; S16878; A36666  
R:Thomas, U.; Spelcher, S.A.; Knust, E.  
Development 111, 749-761, 1991  
A:Title: The Drosophila gene Serrate encodes an EGF-like transmembrane protein with a co  
A:Reference number: S16148; MUID:91347903; PMID:1840519  
A:Accession: S16148  
A:Molecule type: mRNA  
A:Residues: 1-1408 <THO1>  
A:Cross-references: UNIPROT:P18168; EMBL:X56811  
R:Thomas, U.  
submitted to the EMBL Data Library, November 1990  
A:Reference number: S16878  
A:Accession: S16878  
A:Molecule type: mRNA  
A:Residues: 1-1351, T, 1353-1408 <THO2>  
A:Cross-references: EMBL:X56811; NID:G8563; PID:G8564  
R:Fleming, R.J.; Scottgale, T.N.; Diederich, R.J.; Artavanis-Tsakonas, S.  
Genes Dev. 4, 2188-2201, 1990  
A:Title: The gene Serrate encodes a putative EGF-like transmembrane protein essential fo  
A:Reference number: A36666; MUID:91099666; PMID:2125287  
A:Accession: A36666  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-15, 20-26, A, 28-1408 <PLE>  
A:Cross-references: GB:M3759; NID:G158605; PID:G158606  
C:Genetics:  
A:Gene: FlyBase:Ser  
A:Cross-references: FlyBase:FBgn0004197  
C:Keywords: glycoprotein; transmembrane protein  
F:1-84/Domain: signal sequence #status predicted <SIG>  
F:85-1408/Product: gene serrate protein #status predicted <MAT>  
F:85-1221/Domain: extracellular #status predicted <EXT>  
F:1283-316/Domain: EGF homology <EG01>  
F:319-348/Domain: EGF homology <EG02>  
F:335-388/Domain: EGF homology <EG03>  
F:395-488/Domain: EGF homology #status atypical <EG04>  
F:495-526/Domain: EGF homology <EG05>  
F:533-608/Domain: EGF homology #status atypical <EG06>  
F:615-645/Domain: EGF homology <EG07>  
F:652-683/Domain: EGF homology <EG08>  
F:690-720/Domain: EGF homology <EG09>  
F:727-796/Domain: EGF homology #status atypical <EG10>  
F:803-834/Domain: EGF homology <EG11>  
F:841-876/Domain: EGF homology <EG12>  
F:883-914/Domain: EGF homology <EG13>  
F:921-952/Domain: EGF homology <EG14>  
F:997-1060/Region: cysteine-rich  
F:1122-1246/Domain: transmembrane #status predicted <TM1>  
F:1247-1408/Domain: intracellular #status predicted <INT>  
F:1152,196,247,331,412,452,558,739,965,977,1004,1030,1150/Binding site: carbohydrate (Asn

Query Match 88.9%; Score 32; DB 2; Length 1408;  
Best Local Similarity 100.0%; Pred. No. 4e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGPC 5  
|||||

Db 496 AGPC 500

```

RESULT 35
T14602
variant-specific surface protein - malaria parasite (Plasmodium falciparum) (fragments)
C:Species: Plasmodium falciparum
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jul-2004
C:Accession: T14602
R:Voess, T.S.; Felger, I.; Weiss, N.; Beck, H.P.
submitted to the EMBL Data Library, February 1998
A:Description: Identification of a conserved 5' flanking region of Plasmodium falciparum
A:Reference number: Z18158
A:Accession: T14602
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2135 <VOS>
A:Cross-references: UNIPROT:O61077; EMBL:AF050740; NID:G2944094; PID:G2944095; PIDN:AAOC
C:Genetics:
A:Gene: varph17

Query Match      88.9%; Score 32; DB 2; Length 2135;
Best Local Similarity 100.0%; Pred. No. 5.7e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy      1 AGPC 5
Db      897 AGPC 901

RESULT 36
S29216
neurotoxin Tx2 - spider (Phonetrilia nigriverter)
C:Species: Phonetrilia nigriverter
C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 09-Jul-2004
C:Accession: S29216
R:do Nascimento Cordeiro, M.; Ribeiro Diniz, C.; do Carmo Valentim, A.; von Bickstedt, V
FEBS Lett. 310, 153-156, 1992
A:Title: The purification and amino acid sequences of four Tx2 neurotoxins from the veng
A:Reference number: S29214; MUID:93011905; PMID:1397265
A:Accession: S29216
A:Status: preliminary
A:Molecule type: protein
A:Residues: 1-48 <COR>
A:Cross-references: UNIPROT:P29425
C:Superfamily: curratotoxin

Query Match      86.1%; Score 31; DB 2; Length 48;
Best Local Similarity 80.0%; Pred. No. 33;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy      2 GPCV 6
Db      26 GPCV 30

RESULT 37
S29215
neurotoxin Tx2 - spider (Phonetrilia nigriverter)
C:Species: Phonetrilia nigriverter
C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 09-Jul-2004
C:Accession: S29215; B39305
R:do Nascimento Cordeiro, M.; Ribeiro Diniz, C.; do Carmo Valentim, A.; von Bickstedt, V
FEBS Lett. 310, 153-156, 1992
A:Title: The purification and amino acid sequences of four Tx2 neurotoxins from the veng
A:Reference number: S29214; MUID:93011905; PMID:1397265
A:Accession: S29215
A:Status: preliminary
A:Molecule type: protein
A:Residues: 1-49 <COR>
A:Cross-references: UNIPROT:P29424
R:Rezende Jr., L.; Cordeiro, M.N.; Oliveira, E.B.; Diniz, C.R.
Toxicon 29, 1225-1233, 1991
A:Title: Isolation of neurotoxic peptides from the venom of the 'armed' spider Phonetrilia
A:Reference number: A59305; MUID:92196803; PMID:1801316
A:Accession: B39305

```

```

A:Status: preliminary
A:Molecule type: protein
A:Residues: 1-11 <REZ>
C:Superfamily: curratotoxin
C:Keywords: neurotoxin; venom

Query Match      86.1%; Score 31; DB 2; Length 49;
Best Local Similarity 80.0%; Pred. No. 34;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy      2 GPCV 6
Db      26 GPCV 30

RESULT 38
T20083
hypothetical protein C50B6.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T20083
R:percy, C.
submitted to the EMBL Data Library, October 1996
A:Reference number: Z19222
A:Accession: T20083
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-291 <WIL>
A:Cross-references: UNIPROT:O17666; EMBL:Z81050; PIDN:CAB02849.1; GSPDB:GN00023; CESP:CS
A:Experimental source: clone C50B6
C:Genetics:
A:Gene: CESP:C50B6.4
A:Map position: 5
A:introns: 47/3

Query Match      86.1%; Score 31; DB 2; Length 291;
Best Local Similarity 80.0%; Pred. No. 1.6e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy      2 GPCV 6
Db      127 GPCV 131

RESULT 39
T26576
hypothetical protein Y2H9A.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T26576
R:Wallis, J.
submitted to the EMBL Data Library, December 1998
A:Reference number: Z20237
A:Accession: T26576
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-291 <WIL>
A:Cross-references: UNIPROT:Q9U2P5; EMBL:AL021448; PIDN:CAA16275.1; GSPDB:GN00023; CESP:
A:Experimental source: clone Y2H9A
C:Genetics:
A:Gene: CESP:Y2H9A.3
A:Map position: 5
A:introns: 47/3

Query Match      86.1%; Score 31; DB 2; Length 291;
Best Local Similarity 80.0%; Pred. No. 1.6e+02;
Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy      2 GPCV 6
Db      127 GPCV 131

```



## RESULT 40

B44282

retrovirus-related env polyprotein pseudogene - human

N/Alternate names: coat polyprotein

N/Contains: coat protein gp70; coat protein p20E

C/Species: Homo sapiens (man)

C/Date: 17-Feb-1994 #sequence\_revision 04-Jan-1996 #text\_change 09-Jul-2004

C/Accession: B44282

R/Hirose, Y.; Takamatsu, M.; Harada, F.

Virology 192, 52-61, 1993

A/Title: Presence of env genes in members of the RVL-H family of human endogenous retro

A/Reference number: A44282; MUID:93297138; PMID:8517031

A/Accession: B44282

A/Status: conceptual translation of pseudogene

A/Molecule type: mRNA

A/Residues: 1-570 &lt;HIR&gt;

C/Cross-references: UNIPROT:Q9N2K0; GB:D10083

C/Genetics:

A/Gene: env

C/Keywords: coat protein; polyprotein; pseudogene

F:1-372/Domain: coat protein gp70 &lt;CGP&gt;

F:370/Region: opal stop codon

F:373-569/Domain: coat protein p20E &lt;CPP&gt;

F:436-448/Region: immunosuppressive peptide

## Query Match

Best Local Similarity 86.1%; Score 31; DB 4; Length 570;

Matches 4; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 2 GPCV 6

Db 454 GPCV 458

Search completed: December 30, 2004, 15:14:59

Job time : 5.21148 secs

***This Page Blank (uspto)***

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 30, 2004, 14:59:21 ; Search time 61.7885 Seconds

(without alignments)  
2729.763 Million cell updates/sec

Title: US-10-719-385-2

Sequence: 1 MIRSRTSVLSFCRSSREL.....PESQEPILQVQAVRMQR 1753

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 150 summaries

Database : PIR 79:\*

1: p1r1:\*

2: p1r2:\*

3: p1r3:\*

4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	234	2.6	1628	2 T13682	nucleoporin - fibs
2	215.5	2.4	1655	2 T06075	hypothetical prote
3	214	2.4	1655	2 S47446	nucleoporin Np188
4	192.5	2.1	1647	2 T41267	hypothetical prote
5	192.5	2.1	1683	2 S56811	probable membrane
6	184.5	2.0	2670	2 T37919	GCN1 homolog - fibs
7	175.5	1.9	1073	2 T06074	hypothetical prote
8	175	1.9	2110	2 H96803	unknown protein T5
9	171.5	1.9	2104	2 T38774	myosin-3 heavy cha
10	169	1.9	2698	2 B96671	similar to transla
11	167	1.9	4385	2 T39042	hypothetical prote
12	163.5	1.8	2076	2 T28915	hypothetical prote
13	155	1.7	2491	2 A57036	talin - slime mold
14	154.5	1.7	3227	2 T37964	probable ubiquitin
15	151	1.7	3225	2 T00637	hypothetical prote
16	151	1.7	3225	2 I52300	giantin - human
17	151	1.7	3259	1 A56539	giantin - human
18	151	1.7	4464	2 D87755	protein T21E2.4 (
19	149.5	1.7	1334	2 T19493	hypothetical prote
20	147	1.6	2954	2 T14156	kinesin-related pr
21	146.5	1.6	1868	2 S48938	hypothetical prote
22	145.5	1.6	2848	2 T32550	hypothetical prote
23	144.5	1.6	2594	2 A35774	kinase-related pro
24	144	1.6	2693	2 A40743	IP1 receptor, XTP
25	143	1.6	2549	2 S45340	FKBP-rapamycin-as
26	142.5	1.6	2700	2 D84450	protein F21H1.2 (
27	142	1.6	2549	2 A54837	rapamycin/FKBP12 c
28	141.5	1.6	2114	2 E96505	hypothetical prote
29	141	1.6	4924	2 T50176	probable peptide s

30	140.5	1.6	1826	2 H86502	excituclease ABC s
31	140	1.6	2777	2 D96746	hypothetical prote
32	139.5	1.5	4096	2 A57099	DNA-activated prot
33	138.5	1.5	2672	2 A48126	translation activa
34	138.5	1.5	3744	2 S46715	hypothetical prote
35	138	1.5	2241	2 S09811	hypothetical prote
36	137	1.5	2764	2 T13949	hypothetical prote
37	137	1.5	2802	2 T13947	neurofibromin - fr
38	135.5	1.5	2335	2 T40186	probable phosphati
39	135	1.5	1570	2 T42735	TBP-interacting pr
40	135	1.5	1230	2 AC2012	hypothetical prote
41	134.5	1.5	3788	2 T13960	beige protein homo
42	134	1.5	2048	2 C84609	hypothetical prote
43	134	1.5	2470	2 S57085	1-phosphatidylinos
44	133	1.5	1825	2 T42725	actin binding prot
45	133	1.5	1885	2 T30847	actin binding prot
46	133	1.5	2033	2 T30849	actin binding prot
47	133	1.5	2297	2 AB2494	hypothetical prote
48	132.5	1.5	1067	2 T39449	probable importin
49	132.5	1.5	1837	2 T41023	probable nuclear p
50	132	1.5	851	2 T38173	probable phosphati
51	132	1.5	1649	2 T39938	hypothetical prote
52	131	1.5	1896	2 T01490	hypothetical prote
53	131	1.5	4128	2 JC6306	protein kinase (BC
54	130.5	1.4	1920	2 A53188	pericenturin - mus
55	130	1.4	839	1 S71553	endopeptidase C1p
56	130	1.4	2802	2 T13945	neurofibromin - fr
57	130	1.4	4056	2 H96599	protein F4J16.10
58	129.5	1.4	1796	2 AC1895	serine/threonine k
59	129.5	1.4	1888	2 T39009	hypothetical prote
60	129	1.4	821	2 T40994	hypothetical prote
61	129	1.4	1108	2 T06633	hypothetical prote
62	129	1.4	1696	2 T33617	hypothetical prote
63	129	1.4	1966	2 T32552	hypothetical prote
64	129	1.4	2327	2 T31733	hypothetical prote
65	129	1.4	2761	2 T29285	hypothetical prote
66	128.5	1.4	1826	2 D72130	exinuclease ABC.
67	128.5	1.4	3092	2 S46009	GRase-activating
68	128.5	1.4	3796	2 T18514	lysosomal traflick
69	128	1.4	1232	2 T21018	hypothetical prote
70	128	1.4	1321	2 T42842	P-glycoprotein sis
71	128	1.4	1321	2 T42842	bile salt transpor
72	128	1.4	2276	2 T00076	hypothetical prote
73	127.5	1.4	1018	2 T40253	hypothetical prote
74	127.5	1.4	2149	2 C96695	ribulose bisposph
75	127.5	1.4	2513	2 G96536	hypothetical prote
76	127.5	1.4	3433	1 S28381	utrophin - human
77	127	1.4	549	2 D82483	sensor histidine k
78	127	1.4	2535	2 T04824	hypothetical prote
79	126.5	1.4	1630	2 T40217	hypothetical prote
80	126.5	1.4	2330	1 KRIMV	genome polyprotein
81	126.5	1.4	2671	2 A49873	inositol 1,4,5-tri
82	126.5	1.4	2783	2 T31431	inositol 1,4,5-tri
83	126.5	1.4	4845	2 T31067	BIR repeat contain
84	126	1.4	971	2 T10678	hypothetical prote
85	126	1.4	2701	2 S17996	inositol-triphosp
86	126	1.4	3944	2 T19997	hypothetical prote
87	125.5	1.4	1064	2 A41542	adenylate cyclase
88	125.5	1.4	1254	2 T47141	hypothetical prote
89	125.5	1.4	2092	2 S30026	genome polyprotein
90	125.5	1.4	2149	2 S18676	genome polyprotein
91	125	1.4	1325	2 T42722	male-enhanced anti
92	125	1.4	4151	2 T13734	groovin gene prote
93	124.5	1.4	1660	2 A84647	hypothetical prote
94	124.5	1.4	3588	2 T40485	suflactin syntheta
95	124	1.4	8281	1 S76330	endopeptidase C1p
96	124	1.4	958	2 T10679	hypothetical prote
97	124	1.4	1089	2 S53978	PEB1 protein
98	124	1.4	1196	2 S65245	translation elonga
99	124	1.4	3449	2 T01083	hypothetical prote
100	123.5	1.4	3449	2 T07589	disease resistance
101	123.5	1.4	2359	2 T03094	A-kinase anchor pr
102	123	1.4	1101	2 G83637	hypothetical prote

103	123	1.4	1274	2	T25024	hypothetical prote
104	123	1.4	1299	2	AH2090	two-component hybr
105	123	1.4	2331	2	S44054	genome polypeptid
106	123	1.4	2749	1	ACMSIT	inositol 1,4,5-tri
107	122.5	1.4	1823	2	S28974	vitellogenin precu
108	122.5	1.4	1946	2	AC2141	serine/threonine k
109	122.5	1.4	4391	2	A38096	pelecan precursor
110	122.5	1.4	4466	1	S17231	dynein beta heavy
111	122	1.4	710	2	AH1109	anaerobic ribonuc
112	122	1.4	768	2	HE5013	yea protein - Ebc
113	122	1.4	1180	2	A11939	two-component hybr
114	122	1.4	1607	2	T03022	MAP kinase kinase
115	122	1.4	1806	2	T23298	hypothetical prote
116	122	1.4	1953	2	T40642	probable helicase
117	122	1.4	2812	2	T43271	phosphotidylinosit
118	122	1.4	3738	2	T05501	hypothetical prote
119	122	1.4	6486	2	T31076	hypoclidine synthet
120	121.5	1.3	930	2	T00403	hypothetical prote
121	121.5	1.3	1018	2	T18249	conserved hypothet
122	121.5	1.3	1744	2	F86161	Flu03_10 protein -
123	121	1.3	729	2	A81038	hypothetical prote
124	121	1.3	742	2	D85882	hypothetical prote
125	121	1.3	790	2	T19040	hypothetical prote
126	121	1.3	1577	2	T15851	hypothetical prote
127	120.5	1.3	1043	2	G72073	hypothetical prote
128	120.5	1.3	1383	2	G96723	hypothetical prote
129	120.5	1.3	2493	2	T14330	calcineurin inhibi
130	120.5	1.3	2617	2	T40540	hypothetical prote
131	120.5	1.3	2617	2	A82136	peptide synthetase
132	120.5	1.3	3795	2	T00831	hypothetical prote
133	120.5	1.3	4910	2	S64942	probable membrane
134	120	1.3	639	2	F82382	methy1-accepting c
135	120	1.3	1420	2	S54471	probable membrane
136	120	1.3	1755	2	T51532	hypothetical prote
137	120	1.3	2337	2	T40577	probable phosphati
138	120	1.3	2354	2	T13288	mel-41 protein - f
139	119.5	1.3	1033	2	F81595	hypothetical prote
140	119.5	1.3	1043	2	G86550	hypothetical prote
141	119.5	1.3	1207	2	T19041	rae GTPase-activat
142	119.5	1.3	1228	2	S59681	probable membrane
143	119.5	1.3	1422	2	B71437	probable resistance
144	119.5	1.3	1493	2	H71445	hypothetical prote
145	119.5	1.3	1574	1	A38454	myosin MYO2 - Yeas
146	119.5	1.3	1769	2	S53378	probable membrane
147	119	1.3	714	2	AF2458	hypothetical prote
148	119	1.3	2713	2	A55713	histocol 1,4,5-tri
149	119	1.3	3788	2	T30851	lysosomal traffick
150	118.5	1.3	956	2	T19046	rae GTPase-activat

## ALIGNMENTS

RESULT 1  
 T43682  
 nucleoporin - fission yeast (Schizosaccharomyces pombe)  
 C/Species: Schizosaccharomyces pombe  
 C/Date: 21-Jan-2000 #sequence\_revision 21-Jan-2000 #text\_change 21-Jul-2000  
 C/Accession: T43682  
 R/Whalen, W.A.; Yoon, J.H.; Shen, R.; Dhar, R.  
 Genetice 152, 827-838, 1999  
 A/Title: Regulation of mRNA export by nutritional status in fission yeast.  
 A/Reference number: 222629; MUID:99318821; PMID:10388805  
 A/Accession: T43682  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: DNA  
 A/Residues: 1-1628 <MHA>  
 A/Cross-references: EMBL:AF05035; NID:95478223; PIDN:AAD3830.1; PID:95478224  
 A/Experimental source: strain 972  
 C/Genetics:  
 A/Gene: nup184  
 A/Introns: 6/2

Query Match	2.6%; Score 234; DB 2; Length 1628;
Best Local Similarity	19.2%; Pred. No. 3, 1e-07;
Matches 345; Conservative 265; Mismatches 713; Indels 478; Gaps 76;	
Qy	34 SQIBAEINKWRRLLEG---LSYKPP---SPSSAEKRYANKDVAPLKEIGLRISKFL 86
Db	43 SPVYGLTNSN-KTTEBEAKLSYDSSHSISKQLSKSVK-----KISDIT 86
Qy	87 GLDEQSVOLL---QCTLOEDYKGRTRDSVKTYLODEQSQALIKADYYEERTC--IL 141
Db	87 GYNAQVAVYVLVHNYQVELNTQYPSQLDN-DSVLAQEEFOR-----YYAEIISCKVKL 137
Qy	142 RCVLHLITFYODERHPRVEVADCVDKLEKELVSKYQ-----179
Db	138 AFLQACTDADSKKH-----KMATRLIYISICTOARGSEMAQSTPISFCVRII 186
Qy	180 -QFEELYKTAPEWETHGNLITERQVSRWFVQCIREQSMLEIIFLYAVAFEMAPSDLLV 238
Db	187 DYLSKMTSQAPAPASLFTENG---EBAISQWYFFHFNLLQLLRVIFL-----STYSLIV 236
Qy	239 LTRKFKEQGSGKQTNHLYDE-----TMDPVDRIGYFSAIIVEGMDIESLHKALDD 293
Db	237 CNSEMAISWFCMKKTYLHDQEFMHLIDITGSMCKEITNVALIISINFTSLKQVLSF 296
Qy	294 RRELHOPADGLICQDWDCLMTRFGDIPHHAPVLLAWALLRHTL--NPE-----ETSSVV 346
Db	297 KDNPSFPMLSGNTIISLMDITQLSNDISGAASVLTGIALHLNSPDIPILONSVV 356
Qy	347 -----RKIGSTAIQLNVFOYLRILLOSLSAGGNDCTTSTACMCVYGLLSFVLT 394
Db	357 SSKILQNPONSFOALLIAALKYDFLIHRIISL-----EDDVIYIGYSKIMATLF 408
Qy	395 SLELHTGNQODIIDTCEVLADSPLELPMGERTSGLGIIIDVSGMPPHLLSPILLQ 454
Db	409 SSAVSIVKESDSTMLCATTLFKTPQVQLP---ENNDSTRLMPPARAPPEVYSGVYL 465
Qy	445 LRALVSGKSTAKKVVSPFLDKMSFYNELYK--HKPHDVISHEGDT---LMRRQPKLLYPL 509
Db	466 LIPFACLTQKQVLSSELHMTTFQSLPSGFAKYEIIPBNVYGNMLIELQESLHDSY 525
Qy	510 G---GQTNLRIPOGTVCQVN-LDDRAVLVMEVSYSSWTLETCIEMLHNVSTADYIQH 565
Db	526 GFPEPPNARSLPKQTRIRIVSVTPPVVMDLNTYSIMEAVGISLN---YIVRGLNSH 582
Qy	566 QQRKPIIIDLVHKVISTDLSIADCLPITSR-----LYMLLQRLTVIVSPVDVIAS 617
Db	583 KSFVLTVLSSSVPLPQTDVSGACELVHLASBGDELDFINVIDLDLYFLSLVIEDAD 642
Qy	618 ---CVNCLTVL-----AARNPAKVTDLRHGTGFLPFVAPVSSLQWISAEQNNAGGYG 668
Db	643 YQICVSSLRLLREFTRFAVD---VW-----AVYTRSLVCVSGSEKGIS---LE 684
Qy	669 NLAMNSEQPOGEIGVTIAPLRLLT-----TLVKGQIGSISQSIGVPCVMVLKEML 719
Db	685 DVIFDYESINGVVDFTLAFDLYEILLDNCSISTSVVDDEPSIRLKTDFVYKAMRFLCEVF 744
Qy	720 PSYHKRWYNSHGVREOIGCLILELHAILMLCHETDHSHT-----761
Db	745 ANYLDMKRIATIIQOYQGHAFASLITKLNLVNTGIEIFNPKTYYNKKTLPRLRELSHYIQ 804
Qy	762 -----PSILOFLICSLAYTEAGQYVININGIVDTIDWMAAOPRSDGAEGQ 808
Db	805 RFLVQDSNRNYLHPLHSVMDLINLTYDIPSTI-----SSPRAKAK-- 846
Qy	809 GQGQLIKITYKLAFSVNNVY-----RLRPPSNVSPLEQALSQGHAGNNLIAYLAKY 862
Db	847 ---MMLLIS-----SFCAMKTLICLRGFLNLRP-----SELRELFERSRDPFN----- 886
Qy	863 IYHKHDPALPRL-----ATQLKRLATVAPM-----SYAACGNDAAARDAFLTRLOS 911
Db	887 -----CPRLLCCIAPIQLLSAL-TLAPWSESTSLAAYMNSIDIYGRVCIOQLTN 938
Qy	912 KIEDMRIKWILEFLTVAVE-TQPGLELFLNLEVKDGSQSGSKFSLG-MWSCILHAV--- 966

```

Db      939 PIGSTNIEGSAWKLTSSIMKGOQGLAVLFS-----GKFEPLDRMKSLHNVDVQ 989
Qy      967 -----LELISQOQDRWCPCPLLRALAFALHAIAMQDRDSAMLVLTKEPFENLTS-- 1019
Db      990 LTKSKLSLAEKRLDSFINDILSQVFEFIFISRNFWTASLNLQOENAFNMRIVDAI 1049
Qy      1020 --PLFGTLSPSESEBSILETCLIMKIICLEYIY-----VVKSLDQSLKDT 1066
Db      1050 KLPLTWKLDGLSSVAQADLYILAAHATRTITAIQLHMKSLKNSKSKIIIDPLDQSMKDL 1109
Qy      1067 LK-KFSIEKFPAYMSGVYKSLA-----VHVAE----- 1092
Db      1110 VQNAFTIT---AYDSNIHNAITRAFKHNGDLHISDLNLTGLPFLRGDNFYNIKLAKN 1166
Qy      1093 ----TESSCC-----TSLEBYO-MVSAAMRMILTI---IATTHADIMLTDTSV 1132
Db      1167 MLNTEBTSFKSIMMSANENI,SLIDQALRLRMSIFICAFVEFVVEDATLPTLELKIM 1226
Qy      1133 R---ROLFLDVLID-----GTRKALLVPASVNCRL---GSMKCTLLILLRQWK--- 1175
Db      1227 KWLKSLAEEDTIDVNVVOELSAERPAIVFRI,SOOTLAPISNEVKEHLQSTILLTWRKAIT 1286
Qy      1176 -RELGSYDEILG-----PLTEILEGYLQADQQLM--EKTRAKVFSAFIT-VLQMKEMKV 1225
Db      1287 TTKRSIYEDSNGEMAYRRPLHLVYNTL---NRLISEKENLISVGFVSGLDQCHRK 1343
Qy      1226 SDIPQYQVLVNCETLOEVIYALPDQTRISLALGSATEDKDSMETDDCSRRRDQDG 1285
Db      1344 SOL--FEKAVINPTIEVYGDIVLL-----NSLHK----- 1370
Qy      1286 VCVLGLLAKEL--CEVDEGDSWLQVTRRLPILPTLLTLEVLARKMKNHPTEATLHL 1343
Db      1371 -CYNSHLIRLOSLYISYINDSF-SVDNCL-----RIFSMHSLVLDGQYFPDAALSF 1423
Qy      1344 LLTLARTQCATAVAGAGITQISCLPLLSYVQLSTNGTAQTPSARSLDAPSPGVYRL 1403
Db      1424 LLICSSPAGAEQIVMGFFYSIMESPLST-ALSTGGLGIDGS---SIGYKIN---IRG 1475
Qy      1404 SMSLMEOULTLRNPLPEALDFVGHQOERTLOCLNVRTVQ-----SLACL 1450
Db      1476 ILPLLFIVKFLGRINNDKREFY-----LLAFPOQVYLLMWCOPPSISLASI 1525
Qy      1451 EEAHTVGTQLQLSNPKMKNHFLPOLMRDIOVNLGVCAQCTSLHSRKNLGHYLNQKN 1510
Db      1526 DES---FMVLPLDQLQGFN---PALQETR-----LAELKTEMLGISDYILISHPNFLS 1573
Qy      1511 GDGLPSAVAGVQRPPSAAAPSSSKQPADTASECOALHTVQYGLKLISKTLAAR 1570
Db      1574 SLAIPALYEGEVNIDVIGIKELSDQLTKETITSPAKIIRIQ-----LDLRLALLE 1627
Qy      1571 H 1571
Db      1628 H 1628

RESULT 2
T06075
hypothetical protein T9A14.40 - Arabidopsis thaliana
C/Species: Arabidopsis thaliana (mouse-ear cress)
C/Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 09-Jul-2004
R/By: M. Murphy, G. Ridley, P. Hudson, S. Bancroft, I. Newes, H.W. Meyer, K.F.X
submitted to the Protein Sequence Database, March 1999
A/Accession: T06075
A/Reference number: 215184
A/Molecule type: DNA
A/Residues: 1-561 <BEV>
A/Cross-references: UNIPROT:Q97014, EMBL:AL035656, GSPPB:GN00062, ATSP:T9A14.40
A/Experimental source: cultivar Columbia, BAC clone T9A14
A/Genes: ATSP:T9A14.40
A/Map position: 4

```

```

A/Intons: 9/3; 45/3; 64/1; 104/2; 149/3; 202/2; 258/3; 291/3; 317/3; 337/3; 390/2; 457/-
C/Superfamily: Arabidopsis thaliana hypothetical protein T9A14.40

Query Match      2.4%; Score 215.5; DB 2; Length 561;
Best Local Similarity 22.2%; Pred. No. 1e-06;
Matches 132; Conservative 99; Mismatches 244; Indels 119; Gaps 24;

Qy      86 LGDEBSVOLLOCYQEDRGTRDSYKTVLQDE-----RQSQALIL 127
Db      9  MNLDEISYILVERSMQOEY-GTTDSVAQELTQEFIMSGFLHNOCFVWVWESCEAFLL 67
Qy      128 KIADYVEERTCILRCVLLHLYFQDERHPRYEVADCVKLEKELVSKYRQGFELKYT 187
Db      68 TLISLFS--NVILEDASVLTAPVPTKYKAVSHNCI-----ICSVAPRESSIKE 119
Qy      188 EAPWETHGNLMTERQVSRWFVQCLRQSMLEIFLYAYFEMABDVLVTMFEQOG 247
Db      120 EAVVLISDG---LER-----ROSSVLEDLISCF-----PKMGLS----- 153
Qy      248 FGSRQTRHLVDETMDPFVDRIQVFSALILVEGWDISLHKCALDDBRELHQFAQDGL-- 305
Db      154 -GSYNSKLAVSVAQASACRVQIQMLILITIDMEN-----LLQVHVDGVF 201
Qy      306 ---IC-----QDMDCIM--LTFGDI,PHHAPVLLAVAL---LKHITNPEETSSVVRKIG 350
Db      202 RSGTCVPSYDVQEMDATISLNTSEVNAAGPLVLAIVFLCLISLPGKEESPFLMIDID 261
Qy      351 GTAIQLNVFQ-----YTRLLQSLASGANDCTSTACVCYGLSFLVLSLEHTL---G 402
Db      262 HVSYVHOAFERASLSYFELIQSLMLNDFDGPISGHSVVRTPISAFASYEIMQLQEDG 321
Qy      403 NQODIITDACEVL-ADSLBELFWGTE-----PTSGIILDSVCGMFPHLSPDLLRL 456
Db      322 TLELIDILISKVGGESLCCQFMDKSFVDGPRCLLPLES---EPFPSAFIRILS 378
Qy      457 ALVSGSTAKVYSFLDKMSFYNELYKHPDIVSHR--DGLMRRQTPKLYELAGQTN 514
Db      379 SLSEGSMPAEVCVYFLDKSGVSTLF-----DITSDPADASQVLETSRHL-IPGLEG 432
Qy      515 LRITQGVQVM--LDRAVLVWRWYSYSSWTLFTGCEIMLHVAVSADVYIQCQRKPI 572
Db      433 LVIPSNTRGRILVRISNTVLYRWREYSLGIIYLIRLANKLYGNRRAP-----VT 485
Qy      573 IDLVHAKISYDLSADCLPITSRIYMLQRLTVISPP---VDVIASCVNCLT 623
Db      486 DELRRVVTFRKAVCSFLNLSHFYQESTYVNGKMSDVRVVDITCNVSRLT 539

RESULT 3
S47446
nucleoporin Np188 - yeast (Saccharomyces cerevisiae)
N/Alternate names: protein YML103c
C/Species: Saccharomyces cerevisiae
C/Date: 02-Dec-1994 #sequence_revision 02-Dec-1994 #text_change 09-Jul-2004
C/Accession: S47446; S61617
R/Barrell, B.G.
submitted to the EMBL Data Library, August 1994
A/Reference number: S47445
A/Accession: S47446
A/Molecule type: DNA
A/Residues: 1-1655 <BAR>
A/Cross-references: UNIPROT:P52593, EMBL:X80835, NID:G530339, PID:G530341, MIPS:YML103C
R/Zabel, U.; Doye, V.; Tekotte, H.; Wepl, R.; Hurt, E.C.
submitted to the EMBL Data Library, August 1995
A/Reference number: S61617
A/Accession: S61617
A/Molecule type: DNA
A/Residues: 1-1655 <ZAB>
A/Cross-references: EMBL:X90580, NID:G1150597, PID:G1150598
A/Genes: SGD:NUP188
A/Cross-references: SGD:S0004571, MIPS:YML103C
A/Map position: 131

```

C/Superfamily: Saccharomyces cerevisiae nucleoporin Np188  
 C/Keywords: transmembrane protein  
 F/184-200/Domain: transmembrane #status predicted <TM1>  
 F/257-273/Domain: transmembrane #status predicted <TM2>  
 F/376-394/Domain: transmembrane #status predicted <TM3>  
 F/436-448/Domain: transmembrane #status predicted <TM4>  
 F/1336-1352/Domain: transmembrane #status predicted <TM5>  
 F/1499-1515/Domain: transmembrane #status predicted <TM6>

Query Match 2.4% Score 214; DB 2; Length 1655;  
 Best Local Similarity 18.0% Pred. No. 7.8e-06;  
 Matches 350; Conservative 291; Mismatches 695; Indels 610; Gaps 85;

```

QY 31 RELSGLAEALNKHWRRLLEGSLYYKPPSSAEKVKAKNDVASPLK-----76
DB 35 KQLAQIRQLKANTNLLESANTIRQNTSSGDHKKLRSTIANLLQINVDNDPPFAOSD 94
QY 77 -----ELGLRIKFLGDEQSY-----QLLQCY- 100
DB 95 LSHAVEFPMSERSRLHIVSLVNPDPIDLETYSFIDRPNVVGKLISIVQNVDI 154
QY 101 -----LQEDYRGRTDSVKTVALQDERQSAALLKTLADYYEERTCLNCVHL-----148
DB 155 ITASSLADHYNDQDMFTIV-----SLVQLKFS-----LKFTLQILQILNMT 199
QY 149 -----TYFODERHPYRVEADCVDLKEL-----VSKYR-QQFEELYKTEAPTW 192
DB 200 LNTKVPDIYVQWFLQYQNGF-VFCRNINSTKSIDTSSLQLYKFNQDL-----250
QY 193 ETHGNLTERQVSWFPOCLREQSMLEITFL-----YYAFEMAPSDLLVLTKEKQGF 248
DB 251 -----SYLSETLISR-----ISSLFTITITLIGLNTSIAGFDI-QSPLYMDTEFPDVS 300
QY 249 G-SHQTRHLVDETMDFVDRIGES-----ALLIVEGMDIESLHCALDRREH 298
DB 301 ALENDVATNIYNE--DPIFPHMHTYSWSPILYRRALQSSSEFSDSITFAL-----351
QY 299 QFAODGLIQDMCLMLTFGDIPIHAPVLLAMALLRHTLN-----PEBSTSVKRTGTAT 354
DB 352 -FASHSHVLQCLNLTSEILSEFDPYTYTVYVF--LEFSLNPITASTSRVPAKTIISKAP 408
QY 355 QLVNFOYLTRLQSLGAGNDCTSTACMCVYGLSLVLSLHNTLTGNODIITDACEV 414
DB 409 EGFLENLT-----ND-----TEKKLSITIKAPL 434
QY 415 LADPSLELFWGTEPTSGLIIDSVOGMEPHLLSPILQLRALVSGSKTAKKVSFLDK 474
DB 435 LNSLILPLINMLALIDTEFANPELKDIC-----SFAVT 466
QY 475 MSFPLEYKAKPHDIVSHEDGT-----LMROTPLYLPLGGQTN-----LRIPQGTVG 523
DB 467 KSLINDLDYDLADITINSSSSDIIVPDLIELKSDLLVAPLENENSCILSIPIKSTKG 526
QY 524 QVWL-----DDBAYLVMEYSYSWTLTFCHEIMLHV-----STADVI 563
DB 527 KILFTIKQQQQQQQQOQOQPPITSNLIIFLYKENGMSLVGRILQNLHSTYMEKGTQDL 566
QY 564 QHCORVAPIIDLVKVISTDLSADCLLPITSRIYMLQRLTWTYISPPVIVASCVCNLT 623
DB 587 QH-ELMISITIKLVNVDPKSTIEKS-----SEILSVLNSLDTSASTINGAS 633
QY 624 VLA-----KHPAKWTLR-----TGFLLPVAVPVSL-----SOMISABGNAGGYG 668
DB 634 IIVVIFEIFEISLQKDYTSIVOCCEFTMLTPYVHLVSVSYLNKSDLDLKYKT--GLS 691
QY 669 NLAMNS-EOPGEGEVITAFRLITTLVKQGL-----STQSG-GLVPCVMVAKEMLP 720
DB 692 NMLGSELSTGDTTFTIQLIKLVTRISLSLKNIHISKRSKIDITINLLIHAHIEE 751
QY 721 SYHKRVNSHGVBOIG-----CLLELHAHLN--CHETD--LHSGHTPSLOFLCICSL 772
DB 752 SYNMKKNPNFLQKELAFHLTLIFVDVLDHVPFTINPHQKQLLISSANKLQDLPLTPMD 811

```

```

QY 773 AYTEAGOTVINIGIGVDITDMMAPRSDGAGGQGLIKTVLAFSVTNVIRLK 832
DB 812 SIDLAPETLVNLIISPLATTTKILG-----DKILGLYKVMNNSKTLCTLLAIR 862
QY 833 PPSNVSP--LEQALQGHAGNNLAV--LAKYIYKHPALPRLAIOQLKRLATV--A 886
DB 863 GSNRDLKPSNLKFLF-----INSKLDVVTLPYVHF-----VQIELLSYLVEA 910
QY 887 PMS-----VACICGN-DAAAIRDAFIRLOSKIEDRIKIMLEFLTAVET--QPGILEL 939
DB 911 PMWDDYPPFLISFLGEAKSMVFLKEVLSIDLSSPVQNNLRSLYFTFTLLSKQDGLSIL 970
QY 940 FLNLEVQSGDSGKESKSLGWSCLHAVALLELDQQQDRWCPELLRAAIAFLHALMOD 999
DB 971 FLTGQPSANKKINDESI-----DK 990
QY 1000 RDSAMLVLRTPKPFMENTLSPFLGLSPSESEPSILETCALIM-----KIICLEY 1053
DB 991 KSSILVTLQKNSILD-----STPEVS-CKLLETITVLTMTNNSKIFIDPKF 1039
QY 1054 VVKGSLDQSLKOTLKKEPSTIEKRAVSGYKSLAVNAVATEGSSCTSLLEYQMLVAGRM 1113
DB 1040 V--NSLLAKLKOSKFLFOKKENLT-----RDETVSLIKKYLIS--RI 1078
QY 1114 LLIIATTHADIMLTDVVRQFLDVLDTKALLVPASVNCRLGSMKCT-----LLLI 1169
DB 1079 VEIIFALC--INSTDNSSEILNFINQED-----LFELVHNHFQIDGNKTPHEDLNK 1129
QY 1170 LNRQW--KRELGSVDEILGPLTEI-----LEGVLQADQIMETKAKV--FSAP 1215
DB 1130 FKEMPSLEILOSFOKI--PLSRINENENFGYDIPLLDIYLAKDRSWNNEPSKQTNKEBI 1187
QY 1216 T-----VLQKKEMKVSIIPOYSOLVNVCEFLGEEVALPDQRHSLALGSADEKOSMET 1271
DB 1188 TDSALNDQYVNYEISITAKWGAIIITTVK-----KSTVPL-----1222
QY 1272 DDCSRHRDQRBQVCVLGHLAKELCEVDEBDGWSLQV--TRRLPLPLTLTLEVSRLM 1330
DB 1223 -----NDGFVDLVEHFLKLNIDFGSKQMPQTQYLRIBELSFYIIVSPKSGTL 1271
QY 1331 KQNLHTEKTHLLLTATQOGATRVAGAGTOSICLPLS-----VQOLSTNGTACTPS 1386
DB 1272 LKEEKIIELMNKIFITFKSGEIDFIKNIGSKLNNEFYRLBSVYLLELVSSGGRFEL 1331
QY 1387 ASRKSU--DAPSWPVRLSMSMEQLKTRVYNLPALPFGVGHQERTLOCLNAVRT 1443
DB 1332 ISDQLEFEELVPSKGVYILSEILQINK-----CSTRGLSTHTTQIYV--1377
QY 1444 VQSLACLEADHTVGFITQLSNF-----MKWHFHLPLQIMRDI-----1481
DB 1378 -----LE--DNTQDILLILSLFKKITVNPSPKNEFVILASSINEVGLKYLINLYSSAH 1429
QY 1482 -----QVNLGYLQACTSLHSKRMLOHLYQNKNGGGLPBAVAGQVRPPSAAS 1530
DB 1430 LRINDEPLIGQITLFLSELCSIEPIAKL-----NSGLYSVLLE-----1471
QY 1531 AAPSSKQPAADTEASEQOALHTV--QYGLIKLISKTLALRHF-----TPDVCOIL--LDQ 1583
DB 1472 -SPLSVAIQQSDIKQEPSPRLNINWSNGLSTV--LILLSPFGIKVLPETCLFVSYPK 1527
QY 1584 SLDLAEVNFPLALSTFTTFDESVAPSFCTL-----LATVNALNMG--EL-----D 1629
DB 1528 QIKSTIYV-----GDKRLAVSSSLIKETNQLVYLQGMNLNLNQELFIQPKND 1577
QY 1630 KKKEPLTQAVGLSTQABGRITKSLMFMENCFVLLISQAKRYLADPAVHR-----1682
DB 1578 DQENAVELYIGDSEHDKKRSALSKF-----LTHPKYINSRII 1617
QY 1683 -----DKQRMKQELSESLSTILSLSR 1704
DB 1618 PTLLEQQQLDESSRLF-FVKGISR 1642

```

## RESULT 4

T41267

hypothetical protein SPCC290.03c - fission yeast (Schizosaccharomyces pombe)

C/Species: Schizosaccharomyces pombe

C/Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004

C/Accession: T41267

R/Lyne, M.; Rajandrem, M.A.; Barrell, B.G.; Gilbert, H.; Lauber, J.; Duesterhoeft, A. submitted to the EMBL Data Library, January 1999

A/Reference number: Z21982

A/Accession: T41267

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1-1647 (LYN)

A/Cross-references: UNIPROT:P78847; EMBL:AL035260; PDB:CAA22873.1; GSPDB:GN00068; SPDB:

A/Experimental source: strain 972n-, cosmid C290

C/Genetic:

A/Gene: SPDB:SPCC290.03c

A/Map position: 3

A/Introns: 1562/1

Query Match 2.1%; Score 192.5; DB 2; Length 1647;

Best Local Similarity 17.2%; Pred. No. 0.00024;

Matches 343; Conservative 305; Mismatches 673; Indels 677; Gaps 84;

QY 45 RLLEGLSYRK-----PPSSSAEKYKANK-----DVASPLKEGLRISKFL 86  
DB 31 RLILQLESYSKDPLKFLADPANANSRKLESEGEVGVINKVNEQFIQLSLITLSTOL 90  
QY 87 GLDEEGSOVLQCYLQEDYRGTRDSVKTVLQDEROSQALILIKADYYEERTCILRCV-- 144  
DB 91 NLDEIQCASLLQRIEASQNLDRTPVQALY-----PFFLARQLLEGLLES 136  
QY 145 -----LHLTYFQDERHPRYEYVADCVQKLEKEVSKTRQGFELYTEA 189  
DB 137 LTRVWGLKDESDISTALKSYLQSLCENNNLVCTCID-----TPIIDNVSEILKSEA 191  
QY 190 PTWETHGNLWTEROVSM--FVOCLEQSMLEIIFLYANFEMAPSDL-----VLTK 241  
DB 192 -----GGQILGVSEVDPQEFIR--LSHBAVALETTISVILYQLAKVDLFQNSHFESLLV 245  
QY 242 MFKEQSGSRQTNHNLVDETMDPFVDRIGYFSALILVEGMD-----IESLHK 288  
DB 246 MLRR--YDSPKNAVLPLTYAFIDKY-----LEVBYLPDQKVQSRNSVELLQKHQ 297  
QY 289 CALDD-----RRELHQPAQDGLI 306  
DB 298 AIIQSPQDWRSSQFKNILIGIMVTRNLATKQIEKVPFIDYETTIKANANEIIQNGVF 357  
QY 307 CQMDICMLTFGD-----IPHNAVLLMALLRHTLN-----PEETSSV 345  
DB 358 SDMTLLLVPRROSETGEMWAFKRSRSLTVNWSLIRPFIASITFSELSRPAQAVSY 417  
QY 346 VRKIGTAIQLNVQYLRLLQSLASGNDCTSTACVCVGLSFYLSLEHTLGNQ 405  
DB 418 MPDILKRLLEEDRYLTNTTPTSPISPEQIEEYFPPEEYVLLSYTY-----NVS 470  
QY 406 DIITFACEVLADPELPELFMGTEPTSGIITLSDVCGMFRLSLPLQLRALVSGSTA 465  
DB 471 WISD-----FMDIESDMVGFPLTWSMGSPGIIITAFILLASLAKNTTSA 516  
QY 466 KKVVSFLDKMSFVNELYKHKPHVDISHEDGLMRQRPKLILYRGGQNLRIPO--GVYQ 524  
DB 517 SKIT-----ELFSEP-----IPVGHDES 535  
QY 525 VMLDDRAVLVWEXSYSWTLFTCEIEMLLNVSTADVIQ--HCOR-----VKPII 573  
DB 536 LMTISPS-----WSYIFNVFRYIISHIKRYQVTVSSGLARVHTDPSRIDTSAIILQAYI 591  
QY 574 DLYHKVISTDLISD--C-----LPIITSRIYMLQRLTTVISPPVDVIASVCNCLTVIAA 627  
DB 592 LLFSSVRODAQIASCENQDLNPDIATLFEELCECL-----PDSVACIVRALESIAH 645  
QY 628 RNPRAK-----VWTDLRH-----TGFLRFAVHPVS--SLSQMTISABGMNAGVGNL 670

DB 646 LSTGFFNNALMTLDMNFVSVLFDVQGLAPMSIPASIKRSLTKPVTSCGPLNNIRRL 705  
QY 671 LANSQPOGEYGTIAFLILITLVKQ-----LG-STQSGVLPCMFYLKE 717  
DB 706 TVNLE-----KMSFVNLLTSLTRNKSLELVNLPENMIGASYRPPQPYDVYVET 758  
QY 718 MLSYHKRY--NSHGREGQIGCLILELHAILN-----LCHEHTDSSHTEPSIQ----- 765  
DB 759 FVASSTQWRMLRMDGIRLOYAC--LQYMLAVDGLNIDLLYSRLISSKYRDLQNNNL 816  
QY 766 -----FLCISLAYTEAGOT--VINIMGIVDTIDMVAAPRSDGAGQGGQGLLIK 816  
DB 817 HYLITRHPALSLEALVTESVYGLPDLVYGFQLEDD--SVPKT-----IVI 863  
QY 817 TVKLASFVNNVRLKRP--SNVY-----SPLQ 843  
DB 864 TVSASLCILNNVLSLQVLFKNVVPYIAELGISXYILDITISRAYKEVEMTRISIVHL 923  
QY 844 AL---SQGHAGNNLIALV-----AKYIYKHPALPRLAIOCLKLATVAPMSVYACLG 895  
DB 924 ALVGSRHKCFPLSALIELSYLDABGFNNKRPDENKLCSTII--RTANSKRILIFGI 981  
QY 896 NDAAIRDAFLTRLOSKEIDMRKIMLEFLTYAVETQPGILBEL-----N 942  
DB 982 ---RTFESQFLITLSTNDESS-----LILKILLNNLKSGGVYSLALLILGPDISTNVIT 1034  
QY 943 LEVKGSDGSKFSLGMSCLHAYLELIDSQ--ODRYWCPLLRHAA---IAPHALMOD 998  
DB 1035 LRDPQGVYSGRV-----SLMNSILDFIEGRITVNGIEWPVMVQALEIYAFICSC-PL 1087  
QY 999 RDSAMLVATKP-----KFMEN---LTSPLFGTSPSEISEPSILETC---ALINKII 1047  
DB 1088 TSEVTLSTVIRARPELVKMWVGEPIITQOVIQNGGFS--SEESVMSVRCIRSRTOINMML 1146  
QY 1048 CLEIYVYVKSLSLQSLKDTLKKEISIEKRFAPWSGVYSLAVHAYETEGSSCTSLLEYOML 1107  
DB 1147 ATEIHVA--ASVQGN-----KLYNEVYSL---IRTNKHTSTELSGQE-- 1185  
QY 1108 VSAKMLLIATTHADIMHLDTSVVRQLFDVLVDGTKALLVPAVNCRLSGSKCTLL 1167  
DB 1186 -SGFKI-----BEMDI-----LRIDQSIIT--FELPNIPEFNL 1216  
QY 1168 LILLQKRELGSDELGLPTELBEGVLQADQOMKETAQKPSAITYLQMKEMKVS 1227  
DB 1217 NMFTRFDR--GNSD-----FOFDTERVAKIYRLEFEAEMT 1250  
QY 1228 IPQYSLVNVWCETLQEVIALFPQTRHSLSLGSATEDKDSMETDCSRHRDQRDGVC 1287  
DB 1251 -----SLGSAEKTYSWLEQNSKJELAQ----- 1276  
QY 1288 VLGHAKLCEVDEDEG--DSWLQVTRRLPLPTLLTLEVSLEMKONLHTEATLHLL 1345  
DB 1277 -LNTFNHVVLLQDHGCLTAMARLTGIL-----VDCNEVSDIVHEDPIWEVL 1325  
QY 1346 TLRTOGATAVAGAGTQSGICLPLSVYQSLSTNGTQOTSASAKSLDAISWPEVYLSM 1405  
DB 1326 -----RLVLPQVTVNHLGQGT-----V 1343  
QY 1406 SLMEQLKTLRYNPLPALDFVGVGHQERTLOCLNAVTVQSLACLEADHTVGFILQLSN 1465  
DB 1344 SVTSSVLETT---LPHALKRIGALKPEBLG-----KSTY 1374  
QY 1466 FMKEWHFHLQMLKMDIQVNLGYLCOACT-----SLHSRRMLQHYLQNKNGDGLPS 1516  
DB 1375 FMEGIDHIVIGLAKGIQ-----CQGSDESIRENLYGSLISILTVFQKHTSANGDKYDP 1428  
QY 1517 AVAKRVQRPSSASAASSSKQRAADTBASQQLHNVQYGLKILSKITLALAHFTPDV 1576  
DB 1429 VFQDAFOK-----LITNLLTSQFPFDVLTRO--ALYT--NGSCWELSVIILNPLHNVBDI 1480  
QY 1577 COLLDQSLDAEYNFLFALSTPTPTDSEVAFSGTLATVVVVALMMLGELDKKKEPLT 1636

Db 1481 STHL-----YKYLRRNFVSSFLDA-PSRAFSLLSSNKDVLVLSGLEAGQ----- 1526  
 Qy 1637 QAVGLSTQAEGRTRTLKSLMTMENCFYLLISQ-----AMRYLRDPVHPDCKRMK--- 1688  
 Db 1527 --CLLITFAQKMLVSHSVLTIDYIKLAWQMLQCKGGIQLKPYV-----QRLMLQL 1578  
 Qy 1689 -----OELSELSTLLSLSRFRGAPSSPATGVLPSPQKSTSLSKASPE 1735  
 Db 1579 LQIFILVMRYTLKEISNKDLLQSIFFLSRK-----LQDSVQTADQ 1622  
 Qy 1736 SQEPLQLVCAFVHMQR 1753  
 Db 1623 ALSEPNTVKKYVETISR 1640  
  
 RESULT 5  
 S56811  
 probable membrane protein YJL039c - yeast (Saccharomyces cerevisiae)  
 N:Alternate names: hypothetical protein J1216  
 C:Species: Saccharomyces cerevisiae  
 C>Date: 08-Jul-1995 #sequence\_revision 08-Sep-1995 #text\_change 09-Jul-2004  
 C:Accession: S56811  
 R:Poln, T.M.; Aljinovic, G.  
 submitted to the Protein Sequence Database, September 1995  
 A:Reference number: S56793  
 A:Accession: S56811  
 A:Molecule type: DNA  
 A:Residues: 1-1683 <NOV>  
 A:Cross-references: UNIPROT: P47054, EMBL: Z49314, NID: g1008164, PID: g1008165, GSPDB: GN000  
 C:Genetic: A:Gene: SGD: NUP192; MIPS: YJL039c  
 A:Cross-references: SGD: S0003576  
 A:Map position: 10L  
 C:Superfamily: Saccharomyces cerevisiae probable membrane protein YJL039c  
 C:Keywords: transmembrane protein  
  
 Query Match 2.1%; Score 192.5; DB 2; Length 1683;  
 Best Local Similarity 18.2%; Pred. No. 0.00025;  
 Matches 307; Conservative 239; Mismatches 553; Indels 589; Gaps 81;  
  
 Qy 2 IRKSKITSVLSF---C-RSSRELTILLGSALELSQIEALELKNMRLLEGSLYKPP 57  
 Db 108 LRRQYIIQIVSFIVNCFHEDTELYOELIKGAL--VSNI----- 144  
 Qy 58 SPSSAEKVKANKDVASPLKELGLRISKPLGDE-----EOSVLLQCLQEDYGRTRDSV 112  
 Db 145 -----LSAFKFIHTQLSEIKQOINKAQIILENNALFQOYIKRRDPLAREI---DIL 193  
 Qy 113 KTVLQDEROSQALLIKTADYYEERTCILRCVHLTLTYFODERHPYREYADCVDKLEKE 172  
 Db 194 SQLLYG-LVDRKGAIMKNDP-----ILSLHH-----VSELSDN 226  
 Qy 173 --LVSKTRQPFEEELKTEAPFWETHGNIMTERQSRNFVQCLRQSMI-----LEIIFLY 225  
 Db 227 DFFLIYTPAFPHLFASLRVLPPADVYKLHSQ-----FMKDKDSDIYTKVYKVALIFIF 281  
 Qy 226 YAYEMAPSDLLVLTAKFKEQFGSROTRHL-----VDERTMDFVDRIGVFSAIIL 277  
 Db 282 FAYF-----IGCKEDPKRADMTDKTVDDEMTSAVE-LGAIQOLLI 324  
 Qy 278 VEGNDIESLHKALDREELHQAFQDGLICQDMQMLTFGDI---PHNAPVLLAWALL 333  
 Db 325 -----PAADTSIV-EQDKSMELFYDINSLEBHPRIIPKQL 361  
 Qy 334 -----RHTLNPETSSVVRKIG-----GTA-----IQANFYQLTR 364  
 Db 362 DDEKIFSQTNSTNTPASATDNMGRGLMNPSPGMMSTGTALNSMPNNVVEYSTTI 421  
 Qy 365 LL-----QSLASGNDCTSTACMCVYGLSFLVLSL---ELHTLGNQODIITPACSVLA 416  
 Db 422 VLSQGTQEFPLSSPDDVQLTIITDC-----AFLLTKIKDAEEDSLSGEDTLTLDISLKA 476  
 Qy 417 D-----PSLPE--LFWGTEPTSGIGII-LDSVCGMPPHLLSPILLQLLRALVS 460

Db 477 DLRRFLSIVEFYASRREYSCFTMSDKNASNAVFIEMCSRCN--DNLMRSCFLYMLVSLS 534  
 Qy 461 -GKSTAKYVSFL---DKMSFYN-----ELYKPKPHVISHEDGTLMRQTPKLLYPLGG 511  
 Db 535 FGPEMALNVYHYGESSNISIMKNIAOCLSDYTK---ISNFSSLSHKRQ----- 580  
 Qy 512 QTMRLRIPQGVGVMLDRAVYLVRMEYSYSMTLFTCEIMLHVSTADAVIOHQCRVKP 571  
 Db 581 -----QPSSETHN-----DIDSTAAVLEGBLANEA 605  
 Qy 572 II---DIVHKVISTDISADCLLPITGRYIM-LLQRLTVIISPVDVIASCNCLTVLAA 627  
 Db 606 VIFLSLLTLVGSVTYQVDEVDKSSLSKVSVDVLFEPFTKINTPLVGAAFKVISLVLPKLE 665  
 Qy 628 RNPAAWTDLRHGFLPFPVAPVSSLSQMSAAGMAAGGNILMNEQOQGEVGTIAF 687  
 Db 666 SSRKTFWS-----FLDSLIFKSSSLN--YSSBSYR-NAFTNVLT-----KISDVLGF 709  
 Qy 688 LRLITTLVKQQLSGQGLVPCVMFVLKEMLPSTYHK--RRY----- 727  
 Db 710 LQLFNLISHSRNNSEYVPPGKLAFPTLGGYRKVGIMPFYDIFNDILAHVDQIVD 769  
 Qy 728 --NSHGVRQI-----GCLILELHAIALNL-----CHETDLSSHSTPSIQFLC 768  
 Db 770 IRNKRAVQLPIKTIYTGLCSPDYSLNLSIPAANLDALVDCENFPNVYQECFAP--- 826  
 Qy 769 ICSLAYTE-AGQVIVINMGIVDTIDVMAAGRSDBAGQGGQGLIKTVKLAFASTNN 827  
 Db 827 IFNYIFTEKIKYKSIFFNVADVGVDSLEIEG-----GKNQAEEL---QLAVKLINK 874  
 Qy 828 VIRLKPPSNVVSPEQLASQHG-----AHGNN-----LIYVAKTYIH 865  
 Db 875 VLDYQ--ETVEBELPVIKKGKGTDFELPKNYELHGARSYDAIFNIPVLVHGLYV-G 931  
 Qy 866 KNDPALPRALIQILKRIA---TVAPMS-----VYACLGNDAAAIRDAFLRLQSKI 913  
 Db 932 VDDQIATLNSLRILAKLSERSNGSVASLSKRNKLITFDSDV-DESAKIKDAFIQLESSI 990  
 Qy 914 ED--WKIKMILEFLTVANETOP---GLIELFLNLVEKQSGSKSEFSG----- 958  
 Db 991 TDAGVALKIKELIDFLTSLNSYNSRTWTISHLLGPOV-----SNVISGPNLAFTISS 1044  
 Qy 959 -----MMSCLHAVLELDSQODRYWCPLLRRAIAFLHALMORDRDSAMLV-LRTK 1010  
 Db 1045 GTSLDLSLISVLASLNSITKONID--YAPMRLATALEIILKCNRPISGLLYSVLIK 1102  
 Qy 1011 PKFWNTLSPFLFGLTSPSETSEPSILETCALIMKTIICLEIYYVVGKS-IDOSIKDTLKK 1069  
 Db 1103 ENFFERIM-----ILDPQVTR-----FTWNGSPEDNSTEEKCN 1137  
 Qy 1070 FSTIEKRA-----YMSGVYSIAVHVAETGSSGCTSL----- 1101  
 Db 1138 F-IESSEVGAFSLAYRNNYQYL-GLFTHKISFGSTSEVLTVYNYLISNTMYSVRLF 1195  
 Qy 1102 -----LEYQMLVSAEMMLIATTHADIMLT-----D 1129  
 Db 1196 SPLDPLNTGNIGCPKELISLFTVNPVPLNBQVTLNKCSGNIYDFHKMENMLRIKRYAE 1255  
 Qy 1130 SVVRQQLFDVLDGTALLLVPASVNCRLGSKCTLL-----LILLRQMKELGS 1180  
 Db 1256 SLHSNSFSLTV---SKQFLKADAVECIRKAKSHFTNIIISHKALELNLSTLHSMVQ---- 1308  
 Qy 1181 VDBELIGLTELIEGVLQADQQLMEKTKAK-VFAFATIVLQMKMKVSDIQYQSLVNLVC 1239  
 Db 1309 -----LVQIITVDSKEPSTRSNFLEVFGLIP---KISD---YIEFNTPFS 1350  
 Qy 1240 ETLQOEVIALFD--QTRHSIATGASATD-----KDSMETDDCSRSR 1278  
 Db 1351 EBLVSLAVLFLFDLYNDRKILTDKGVLDGRLYDLFTKTCIGINSPLSSVALRBDFFILAN 1410  
 Qy 1279 H-----RDQRDG-----VCVLGHLAKELCEVDEDG 1304



Db 1411 HYLRSVLSDQVSEKVLQDLRLGSKVLVEIIMNDVYGEGRVYTGILLDLSLIQLANRS 1470  
Qy 1305 -----DSMLOVTERLPIPLTLTLTLESLMKONLHTE-----ATLHLLTLTLAR 1349  
Db 1471 KENFIDLSMKTTLTLIRSLKNTDNLNSTTHINIDLLYELTAFKATVPLINVAE 1530  
Qy 1350 TOOGATAV 1357  
Db 1531 TRGASAL 1538

RESULT 6  
T37919  
GCN1 homolog - fission yeast (Schizosaccharomyces pombe)  
C:Species: Schizosaccharomyces pombe  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
C:Accession: T37919  
R:Conor, R.; Church, C.M.; Bartell, B.G.; Rajadream, M.A.; Walsh, S.V.; Wood, V.  
submitted to the EMBL Data Library, December 1995  
A:Reference number: 221754  
A:Accession: T37919  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-2670 <CON>  
A:Cross-references: UNIPROT:Q10105; EMBL:Z68198; PIDN:CAA92385.1; GSPDB:GN00066; SPDB:SH  
A:Experimental source: strain 972h-; cosmid c1866  
A:Gene: SPDB:SPAC1866.05c  
A:Map position: 1  
A:Introns: 50/1

Query March 2.04; Score 184.5; DB 2; Length 2670;  
Best Local Similarity 18.64; Pred. No. 0.0019; Matches 697; Indels 699; Gaps 89;  
Matches 380; Conservative 269; Mismatches 697; Indels 699; Gaps 89;

Qy 20 LMTLLGRSALRELQIEAEI-NKGMRLLEGLSYKPPSPSAEK--VKANKDVASPL 75  
Db 797 LMDVLEKSSKLDKMTKQYETKMEAEVAKQSAKKPAKLSKQALVDAQDAEAKI 856  
Qy 76 K-----ELGLRISKPLGLDEBQSVQLQCYLED-----YKGTDSYKTV 115  
Db 857 RSRVNLALSLERGLGIIRSLG---EAVQLAPALWEADIVLLFNHVLKYSSEPLKNTL 912  
Qy 116 LQDRQSOALILKADYVEER-----TCILRCVLHLTYFQDERHPHYREYVADCYDKL 169  
Db 913 AYD-----TFLLTKASGFERLGDRTSYSSLSLIL-----AHTFSVNSSENTEL 958  
Qy 170 EKEIVSKYRQOFELYTEAPTEWTHGNLMTERQVSRWFVQCLREQSMLEIIF--LYYA 227  
Db 959 TKSLTYLTLRFAIEQNY-----FEPQMFACIFPLLYDL 990  
Qy 228 YEFMAPD-----LVLTYMFKEGFGSGRQNR-----HLVDETMDPPVDRIG 270  
Db 991 TFNITNSDEDEAEALQILVTEILFEQALYSASLRMSKLIKSLHLLEIAPTOQENKN 1050  
Qy 271 YFSALIIIEGMDISLHKCALDDRELHQFADGLIQDMDCMLTREGDIPHHAPVLLAW 330  
Db 1051 -SLSLICEG-----LHSTYDEELNL----- 1070

Qy 331 ALHTHTNPEET--SSVVRKIGTAIQLNVQYLTRLQSL-----ASGNDCTTSTA 381  
Db 1071 -LLSNLHRPESISAVIQL--QAFDLSREFPEIKETLELYXDNETNAGIAHQISTQN- 1126  
Qy 382 CMCVGLISFVLTLEHTLGNQ-----QDIIDTACEVLAD----- 417  
Db 1127 -----GLDATERSEFEIQLIFPTQSDYLOQIIIGKSLDILDEFEELQGFPEKELMRTYRE 1181  
Qy 418 ---SLPEL-FWGTPEPISGL-----IILDSVCGMFPH-----LISPLQLL----- 455  
Db 1182 NALPSAPYDEYGIITKETTRDGLRIARBSVAVSFFIYSKYLSSNLLPLLEPLTAGE 1241  
Qy 456 -RALVSGKSTAKKYVFL--DKMSFYN-----ELYKHPHDIVISHEDETLMWRQ 501

Db 1242 AEAQIPVTDASQKVSFKLEAGKLATFQSGAHQVEALMELEOKLNVDSLPTDANRLRE 1301  
Qy 502 TPKLILYLGQTNLRIPQTVGQV--LDDRVLVREYSSMTLPTCEIMLHVST 559  
Db 1302 ATVVLF-----GTVAQHLPSNDPLAV-----MDSILSVLST 1334  
Qy 560 ADVLQHCQVRIIDLVHKVISTDLSIADCLIPTRSIYMLQRLTIVISPPVDVIASCV 619  
Db 1335 PSE-----SVGLAVAVCLPLVKK----- 1353  
Qy 620 NCLTVLAARNPAKWTDLRHTGRLPFVAHPVSSLQMSISAGMAGGYGULMNS---E 675  
Db 1354 -----SLGSKKEYEEL-----LSNKLNNSTSLAD 1377  
Qy 676 QPQGEVGTIAFLRLITTLTKGQSGTQSGIVPCWFVLKEMLSYHKRKYNSHGREQ 735  
Db 1378 QKGAAY-------LAGLVKG-----YGIKAFQDFNLIDSLSELISNRQNA--THRQ 1420  
Qy 736 IGCILTELHAIINLCHETDLHSHSPSLQPLCISLATEAGQTVINIMGIGVDITDMV 795  
Db 1421 VALFAVAFSRILGITYE-----PYLPDL-----LPLILTSFGDNANEVREATMDAVKQI 1470  
Qy 796 MAA-----QPSDGAEGG-----QGOLLIKTV--KLAFS 823  
Db 1471 MSQLSAFGVKLLPTLLDGLNEVYMRSKKASVEILGLMSYAPKQLSVFLPTIIPKLSEV 1530  
Qy 824 VTNNVITLKEPNS-----VPSLEQALQHGHNMLAVAKIYI-H 865  
Db 1531 LITDSHGVNTANKSLRFGDVISNPEIQTLVPLTLALSDCTRYTDALALTKTSFVH 1590  
Qy 866 KHPDLPRLAIQLKRLATVAPMSVYACLGNDAAIADAFITRLQSKIEDMRIKVMILEF 925  
Db 1591 YLDP--PSLL-----VITILKY-----GLRENAQTKQSAKIPIGLMASITPEPN 1634  
Qy 926 LTVAVET-QPGLIELFNLSEVKD-----GS-----DGSKFSL--GMSCLHVALE 968  
Db 1635 LAVYLSIMPLRLREVLI-DVPTPRATAKALGSLIEKKEKFPPLIPELFWLRESCS 1693  
Qy 969 LIDSQ-----QODRYWCPEPLHRAIAFLHLMQDRDSAMVLTKRP 1011  
Db 1694 EVDQGAAGUGSETIAGLGLARLED--VLPEILKNTSSPVPHI---RSEFISLLYLP 1746  
Qy 1012 KFMENITSPFGTSLSP-----SETSEPSILETCALIM-----KIICLEIYVYVKS 1058  
Db 1747 ATGSSRFQPIYLAIRIPIILSGLADDSLVGTASIRAKMIVNNYATSVLLLEPELKG 1806  
Qy 1059 LDQSLKDTLKKFSIEKRPAYMSGVYKSLAVHVAETEGSSCTSLLEYOMLVSAMMLIIA 1118  
Db 1807 FDNAMRIRLSVQVLGDLVPEKLAGINKKALOEDDEEGTHSD-----VSRKALLDIIIG 1859  
Qy 1119 TTHADIMHLDSVVRQLFDVLDGTALLVPAVNVCLALGSMKCTLLILLRQWRRL 1178  
Db 1860 QERHRIILSTLYTRQDI-----AAVVRTPA-----IQTWYAIIV 1893  
Qy 1179 -----GSYDEILGPTLEILEGLADQ-----OLMEKTKAYVSAFITVLO--M 1220  
Db 1894 VNTPRTYREILPITLSIIVSNLSSNDKRTMCKYSGDLIKKAGFVYLQQLPVLQGL 1953  
Qy 1221 KEMKVS DIPQYSQVLVNVCELTQEVIALFPDQTHSLALGSATEDKXMETDSCSRSHR 1280  
Db 1954 ESANGD-----RIGVCIALEB-----LINSATPEQLEIYSDDPVAVVR 1993  
Qy 1281 DQRPQVCVGLHAKELCEVDEGDSWLOVTRRLPIPLITLTLEVLBRKQNLHFPEAT 1340  
Db 1994 ALMDG--DLEVETAEAFDSLOSILGDAVVDVLPOLKLLSESNQESQALSALREI 2049  
Qy 1341 LH-----LTLTL-----ARTQCATVAGAGITQSICTPLSVYQSLSTGTA 1382  
Db 2050 ISRRSITIFVLIITLTKKVSAPNABALSSIAQVATVINKRLPSILNLMESLASTG 2109  
Qy 1383 QTPSASRKSILDAPSPGVYRLSMSLME-----QLIKTLRYNPLPALDPVGVHQRITQL 1438  
Db 2110 DDIVALNGAID--KVNLSVKQEGIQILIMAHFYSF--SESEDF-----RKR----- 2151

QY 1439 NAVRTVQSLACLEADHTVGF-----LQLSNFMKEMHFLPQLMWDIQVNLGYLQACTS 1494  
 Db 2152 -----LFAAEHMLVFQONCKLDYRYRGDWHFHTLTFEDKSQDVVAVAQAQN 2200  
 QY 1495 LHH--RK-----MLQHYLQNNKGG--LPS-AVAQRY-----QR 1524  
 Db 2201 TLVSALRRQQLDLSVSIYVHSLRDVSGQVNLPAFEVAQGVNSILPIFLYGLMGTMDQR 2260  
 QY 1525 PPSAASAPSSSKOPADTEASEQOALMT-----VQYGLKLIS 1563  
 Db 2261 EOSALGADIVLK-----TEPSKMPFYQTGTPLIRIGRFPVEVCKALTYLNTILIS 2315  
 QY 1564 KTLAALRHFTPDVQCILLDQSLDLAEVNFALFALSTFTPTD---SEVAPSGTLATVNV 1620  
 Db 2316 KISTLRPFPLQOLR-----TFKCLGDPSSSEVISAATLALGTLIT 2357  
 QY 1621 ALNMLGELDKKKEPLTQAVGLSTQAEGRITLKSLLMTWENCIFYLLISQAMRYLRDPAVH 1680  
 Db 2358 -----LQRLAPL-----ITELVSGARTPDAGVRKAMNALPAVVSQGNNEASAE 2405  
 QY 1681 PRDKORMQELSELSTLSSLSRYFRGASPSPATVGLPSPQKSTSLSKA--SPESQEP 1739  
 Db 2406 A--IEQLDEISAESSEHMTCAKLY--GALFSH---LPDAQKQLLESKVSLEIQSE 2457  
 QY 1740 LIQTV 1744  
 Db 2458 FSVLI 2462

## RESULT 7

hypochemical protein T9A14.30 - Arabidopsis thaliana  
 C/Species: Arabidopsis thaliana (mouse-ear cress)  
 C/Date: 30-Apr-1999 #sequence\_revision 30-Apr-1999 #text\_change 09-Jul-2004  
 C/Accession: T06074  
 R/Beyan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Mayer, K.F.X  
 Submitted to the Protein Sequence Database, March 1999  
 A/Reference number: 215184  
 A/Accession: T06074  
 A/Molecule type: DNA  
 A/Residues: 1-1073 <BEV>  
 A/Cross-references: UNIPROT:Q9T013; EMBL:AL035656; GSPDB:GN00062; ATSP:T9A14.30  
 A/Experimental source: cultivar Columbia; BAC clone T9A14  
 C/Genetics:  
 A/Gene: ATSP:T9A14.30  
 A/Map position: 4  
 A/Intons: 45/3; 93/3; 129/3; 184/2; 343/1; 451/3; 977/3; 1047/1  
 C/Superfamily: Arabidopsis thaliana hypochemical protein T9A14.30

Query Match 1.9%; Score 175.5; DB 2; Length 1073;  
 Best Local Similarity 17.6%; Pred. No. 0.0018;  
 Matches 213; Conservative 189; Mismatches 416; Indels 393; Gaps 51;  
 QY 694 LVKQGLSTSGQLVPCVMFLKMLPSYHKMRNHSVGEQGLLELHAIIN----- 749  
 Db 3 LVBSGL---ENDVVFALVPSLQYITLASHYWKYKNGMNMKVTLKVIELMKTCLRSKF 59  
 QY 750 -----LCHETDIHS-----SHTPSIQFLICIS-----LAYTEAGQT-VIN 783  
 Db 60 STKLADVLIDLINDASVHSLPFIICITTONLETTQVCAQVLSKLFALAESQLYIIS 119  
 QY 784 IMGIGVDT-----IDVMAAOPRSDGAEQGGQGLIKITVKAAPSVTNNVIRLKPSSNV 838  
 Db 120 NAGFGILNKKPALVIAIFDDEDESDSN-----VQSRKDSASIPDMACKSRLLHTTIL 172  
 QY 839 SPLEQALSGAGGNNLIYLAKEYHKHDPALPRLAIOQLKRLATVAPMSVACIGNDA 898  
 Db 173 QYERADTFPRDRHTDILGLL-----DFLKTLMQEA--GQYA----- 207  
 QY 899 AAIRDAFLTLQSKIEDMRI-KVMILEFLTVAVETQGLIELFLNLEVKD---GSDGSK 954  
 Db 208 -----NMLEEFKASKKLMQEFSDIISQAS-----KINDSTVGSIGKEE 245

QY 955 PS--LGWMSCLHAVLELIDSGQODRYWCPPLHRAIATFLHMLMDRRDSAMLVLRTPK 1012  
 Db 246 ISKLIVYQCOQASVLEIM-----ACNNFLYKGL 274  
 QY 1013 FWEMLTSPPLGT-----LSPSET-----SILETCALINKITCL-- 1049  
 Db 275 FASLSKKPCVETKKTASNGVSPKLTWTADSDPKDIFSKWCDISVD--GIQVSGSLDG 332  
 QY 1050 --EIVYVKG-----SLDQSLKDTLKFSIEKRFAPWMSGYVSLAVHVAETEGSSC--T 1099  
 Db 333 ESEINFQAKRKLEGRDIPFGPKELFQFLVETSF--WEKYKQKTMQDVMLAGDCLFDT 390  
 QY 1100 SLEBYQMLVSAM-----RMLLIATRAHIMHILDSVVRQLF 1137  
 Db 391 QQTQTEIGLIDWDPSEWKTSTTAEMLYMORANSMTLSQSLVHALISVL--IIV 448  
 QY 1138 LDVL-----DQTKALLVPASVNCIRLGSW---KCTLLILLRQMKRELGS 1180  
 Db 449 EDNISVHLNATPVYQSLSESAALERKIPSRVTLISIDKCRKCTTVDSIASIW----- 502  
 QY 1181 VDEILGPLETILEGVYQADQQLME---KTKAKVSAFTVYLQKE---MKVSDIPQYSQ 1233  
 Db 503 -----DAKIVPDILTAQADILSRLLKSAKXKLSVQALVLRNVPGLIKIGSLRHSN 556  
 QY 1234 LVINVCETLQEEVIALF-----DQTRSLALGATEDKDSMETDDCSRRHRDQDGVCL 1289  
 Db 557 AILKKTINLLEVLVLVVGSGSNSNSGGMVTLAKDAEISDAT-----I 603  
 QY 1290 GL-----HLAKELCEVEDGDSMLQVTRRLPIPLTLTLLEVSLRW--KQNLH 1335  
 Db 604 GLPLPLCNFNGNPEVYTLCLTIVDILIRNFIPEFTWFPILIQSLRQHVILIQDQKST 663  
 QY 1336 FTEATLHLLTLARTQCGATAVAGITOSTICPLISVYQLSTNGIAQTPSARSKLDP 1395  
 Db 664 SVSAILKFFLTIAQVHGAQMLNSGFFSTLRALMEF---PDGMSTLVSDEKG--- 715  
 QY 1396 SWPGVRLSMSLMEQLKT-----LRYNPLPE-- 1422  
 Db 716 -----SLEKTEKQIHMGIGLAVVTAMVHSLSGVSAGADIVESVIGYFLEKGY 765  
 QY 1423 -----ALDPVGHQERTLOCINAVRTVQSLACLEADHTVGFILQLSNFMKEMHFLP 1475  
 Db 766 MTSYLLAADFPSS--DDRDKVRLRSQRTWTSIAYLRVETHLLICALASHRSWVKIM 823  
 QY 1476 QL-----WRDIQVN--IGLQCACTSLHSRKLQH-----YLQNNKG-- 1511  
 Db 824 DMDSPURENTIHULAFISGAQRLRSQSHISHLCPVAKKEFPDCKRPSFINTGHWF 883  
 QY 1512 -----DGLPSAVVQVORPPSAASAPSSSKOPADTEASEQOALHTVQGLKILSK 1564  
 Db 884 SLAPLCVCKPRTIVASIS---TALVNRDTEHNGSVQSGFSSVAQIRVAVSLIK 940  
 QY 1565 TLAA-----LNHF-----TPDVQCILLDQSLDL--AEYNFLFALSTFTPTF 1603  
 Db 941 FLCTQAEVGTVAEEVGYDIAHFPDLPEPFIHIGIQOQATIAVAB---LCDNYSKKEI 996  
 QY 1604 DEVAAPSFGTLATVNVALNM-----LGEBD---KKKEPLTQAVGLSTQAE- 1646  
 Db 997 PDEVKKLCMLQOTTEMSLYLELCVQVCRHINPVGGRVDSKDKLVKAAEVHTYLER 1056  
 QY 1647 GRTTLKSLIMF 1657  
 Db 1057 SIDSLSKIAAF 1067

## RESULT 8

H96803  
 unknown protein TSM16.5 [imported] - Arabidopsis thaliana  
 C/Species: Arabidopsis thaliana (mouse-ear cress)  
 C/Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 09-Jul-2004  
 C/Accession: H96803  
 R/theologis, A.; Becker, J.R.; Palm, C.J.; Federpiel, N.A.; Kaul, S.; White, O.; Alonso,



A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Accession: 12104 <SKE>  
 A:Cross-references: UNIPROT:O14157; EMBL:Z98762; PIDD:CA11475.1; GSPDB:GNO0066; SPDB:SH  
 A:Experimental source: strain 972h(-); cosmid c48  
 A:Motif: F.; Nakano, K.; Kiyama, C.; Yamamoto, M.; Mabuchi, I.  
 FEBS Lett. 420, 161-166, 1997  
 A>Title: Identification of Myo3, a second type-II myosin heavy chain in the fission yeast  
 A:Reference number: Z22385; MUID:98119399; PMID:9459302  
 A:Accession: 143276  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Accession: 1-2104 <MOT>  
 A:Residues: 1-2104  
 A:Cross-references: EMBL:AB007633; NID:G2828347; PIDD:BA24579.1; PTD:G2828348  
 R:Bezanilla, M.; Forsburg, S.L.; Pollard, T.D.  
 Mol. Biol. Cell 8, 2893-2705, 1997  
 A>Title: Identification of a second myosin-II in S. pombe: Myo2p is conditionally required  
 A:Reference number: Z22572  
 A:Accession: 143553  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-1132; 'G', 1194-1303; 'G', 1305-1343; 'K', 1345-1419; 'D', 1421-2104 <BEZ>  
 A:Cross-references: EMBL:AF029788; PIDD:AA04615.1  
 C:Genetic:  
 A:Gene: SPAC4A8.05c; myo2  
 A:Map position: 1  
 C:Function:  
 A:Description: may be involved in cytokinesis and stabilization of F-actin cables  
 C:Superfamily: myosin heavy chain; myosin motor domain homology  
 C:Keywords: actin binding; ATP  
 F:92-755/Domain: myosin motor domain homology <MMO>

Query Match 1.9%; Score 171.5; DB 2; Length 2104;  
 Best Local Similarity 16.3%; Pred. No. 0.01; Matches 627; Indels 739; Gaps 89;  
 Matches 323; Conservative 298; Mismatches 627; Indels 739; Gaps 89;  
 100 YLQEDYKGRDSS--VKTIVQDEROSQALIKIADYEEERTCILRCVHLITVFQDERH 156  
 3 YLQKN--GSDNNNIIKKLVDAEKHNA--VQDAFDERTWI-----WIPDSKE 47  
 157 PYRVEYADCVKLEKEVSKYRQOF-----ELYTEAPTEWETGNTLTERQVS 205  
 48 SPYKAW-----IVEDGKRYKRVGLERDGERIVDGFDAKVNPPKEDMDVMAA----- 96  
 206 RWFVQCLREQSMI-----LEIIFLYAYVFEMAPSDLLVL-----TKWFK- 244  
 97 ---LITCLNESPVANNLTQRYEKDLITYSGIFVAVNPPCHLPYGDVVRKTKQSKQFKE 153  
 245 -----BQFGSRQTNHNLVDETMDPFVDRIGYFSAL 275  
 154 TKDHIFGTADAAYRSLLEKRIINGSLVITGSGAGKETETKKVIQ-----YLTSV 202  
 276 ILVEGMDIESLHACALDDRELHOPADGLICQDMC-----LMTTGGDIPHHAPVILAW 330  
 203 TDASTDSQOLKEKLTETNPLEAFGNAQTVRNNSRFGKFIREFSNNGSIVGANLDM 262  
 331 ALNR--HTLNPETS-----SVVRKIGTAIO-----LNVFOYLTLLQSLASGN 374  
 263 YLLEKSVITIPSSNERYHYNYFYQLIRGADSLSEFLDXYVDHYSTLRKGLKHI--NGV 320  
 375 DCTTSTACMCVYGLSLFVLTSLSLTLT-----GNODIIPDAACEVLAD----- 417  
 321 DDKKFKQKLC-FGLRTLGFNNNEIHSFLIILSIHGN-----IEVASRSGGARF 371  
 418 PSLPELFWGTEPTSGGLIILDSVCGMPHLS--PLQLLALVSGSTAKKVSFLDK-- 474  
 372 PSLTQ-----IDOLC-----HLLEIPVDFVNAALHPXSKAGREIVTARTR 413  
 475 -----MSFYNELYKHKPHDVISHEDGTLMBRQTPPLVPLGGQTLRLPQGVQVWL 527  
 414 EGVVHTLQSLAKGLYENRFAHLVRLNQITWYSOSEH-----DGFITG--VL 457  
 528 DDBAYIVRWEYSYSSWT-----LFTCEIEMLHVSTADVIQHCORV 569

458 DINGFEI---FTTNSFEQCLINTNEKLOQFPHHNVFVLEGEETGERIEWDITIDYNDL 514  
 570 KPIILVHKVISTDLSI-----ADCLPI-----TSRIYMLQRLTTVISP----- 611  
 515 QPITDAIER--SRPIGIFSCLEDCEWPMATDATEREKLHLLEPKGSDIRPPKFSSEGF 572  
 612 -----VDVIASVNCITVLAARNPAKWTDLRHNGFLFVNAHPS 651  
 573 VLKHYADVEYDTKMLEKKKODPLNACLALMFKSTN-----S 610  
 652 SLSQMTSAGMNAAGYGNILMNSQOPQGRY-GVTIAFLRLITLVKQGLSGTOSQGLVPC 710  
 611 HVSLSDYDVSNASGNDI-----EKKGIFRTYSQHRRLQSLSLMH-QLEATQPH-FVRC 663  
 711 VNEVLKEMLPSTYKWRYN--SHCVRQIGCL-ILELIH-----AIIWLCHETDLSHSTPS 763  
 664 I-----IPNNIKQPHNLDKSLVHLQRCNGVLEIRIAQTGFPPKLFYTEFRARYGIL 716  
 764 LQFLICSLAYTEAGQTVINIMIGIVDTIDMWAAQPRSDGAGGQGLITKTVLAFS 823  
 717 SQSL--KRGYVAKKATIT-----INELKL----- 740  
 824 VTNNAVIRLKP-----PSNVSPLEQALSOHAGNNLILAVL-----AKTYI- 864  
 741 -PSTVRLBETKVFVFSVLSLED-----RRNALLVYFNSFSARIRGFLTRRLYR 792  
 865 --HKDDPALPRALQILKRLATVAP-----MSVYACIG--NDAAIRDAFLRLQS 911  
 793 FNRQDAAL--LQHNLRQILKLPHPWNLFLHLRPLLTGTOTDYLRLKRLINLQN 850  
 912 KIEDMRKIMILEFLVAVETOPGLLEFLNLEVKDGSQSKFSLGMSCHAVLEID 971  
 851 QLESTR--EVANELLITKE-----RVQLQTN 874  
 972 SQQDRYWCPEPLHRAAIFLALMODRDSAMLVLRTPKFEENLTSLFGLTSPSET 1031  
 875 DLQEBQ-----ALAHKQDILVERANSRVEVHERLSLENQV-----TIADEK-- 917  
 1032 SEPSILETCALIMKICLIRIYVVKSLDQSLKDTLKRS-----TEKRAVMSGY 1083  
 918 -----YEFVLAEKQSIEDLANKQTEISYSDLSSTLEKKSJIKQDE 960  
 1084 KSLAVHVAETEGSSCSLLEYQMLVAMRML-----LIITATADIMHLLDSVVRQ 1135  
 961 QITSSKXKLEKDYINMADYQHSQHLNLEKAINREKUNIRELNEKMLRDLDELKQ 1020  
 1136 LFLD-----VLDTKALLLVPAVNCIRLGSMTLLIL--LRQW 1175  
 1021 RSYDTKVOELREBNASLKDQCRTYESQLASIVSKYSETESELNKAELAVIFQKEITERYR 1080  
 1176 RELGSV--DEILGPLELLEGLVLODQOMETKAKVNSAFITVQOMKEMKSDIPQVS 1232  
 1081 DQHLKAFQNPBETHNINDVSGPLNSDENIY--STSTTSLDKQVELSLHATKENAQLS 1139  
 1233 QLVANVCETLQOEVLIA-----LFDQTHSLATGATEDKDSMET 1271  
 1140 ERKEISEMLEQSIAREEKLARKNSLCOIIEALKYQIQDQETHEISINA-----DNLDL 1194  
 1272 DDCSRHRDQDRGCVLGL-----HLAKELCEVEDGDSMLQVTRRLPLILP 1318  
 1195 KOTNGVLEKNASDFIDFGIKRVEBKISDLNQLQEKERCKVQ----- 1237  
 1319 TLLTLEVLKMKQNHFEATLHLLTLARQOQGTAVAGAGITQISCLPILSVQQLST 1378  
 1238 -----LKKQTEKNSVTOHTL----- 1253  
 1379 NGTAQPSASRSKSLDAPSWGVRL-----SMSLMEQLKTLRYNPLPALDVGVOHER 1433  
 1254 DQNSPHPSFBEKSGSP-----LKRIDGNNDKIDKDKLKT-----ISKSLDLQGLVEE 1304  
 1434 TLQCL-----NAVETVQ--SLACEADHTVGFILQLSNPKMEHFLHP 1475

Db 1305 ELSNLSYSLSKDLSFTDIGHIIPNSIRKLEKSLSTSEPKER-----LANSNSDRP-----SP 1356  
Qy 1476 QLMRDIONVGLVGCACSTSLHSRMLQHYQNKNGD---GLPSAVARVQRP----- 1525  
Db 1357 DIFKDTQ-----AINNSRLL-----LSNPSNDQSGISSLQKLLNPNESNMEFT 1401  
Qy 1526 -----PSAASAPSSSKQPADTEASEQALHTVQYGLKLSLTYLALRHF-----TP 1574  
Db 1402 GLKPLSPKISNLBSS--QPSPSKRS-----GGMELINFDQNSISIP 1443  
Qy 1575 DVCCILLDQ-----SLDLAEYNFLPALSFTTPTPDSEVAPSGFTLLATV-----NVA 1621  
Db 1444 D--PPIVQGRNSVLQTEPEKINLKKKEATKSGILDNKOLSKFSELIQSLKENBELKYL 1501  
Qy 1622 LNMGEIDKKKE--PLTVAGLSTQAE-----GTRILKSLMTMENCPIYLLISQ 1669  
Db 1502 TSNIGSDDKMLDFAPLLDVPNTTNOIKGFVEKAISSKRAIAKLYSASEELF----- 1555  
Qy 1670 AMRYLRDPAVHPRDKQKQKSELSTLLSLSRYPFRGAPSSPATG-----VLPSPPQ 1723  
Db 1556 -----STKRLREITTKERDLHLGLQ---GPSVPTSLKAPLASQLIIPND 1599  
Qy 1724 GKSTSL 1730  
Db 1600 GSITNYS 1606

## RESULT 10

B96671  
similar to translational activator [Imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 09-Jul-2004  
C:Accession: B96671  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federpiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasey, T.H.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Hulzar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719; PMID:11130712  
A:Accession: B96671  
A:Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-2698 <STOP>  
A:Cross-references: UNIPROT:Q9XIR5; GB:AEO05173; NID:gs042415; PIDN:AAD38254.1; GSPDB:GN  
C:Genetics:  
A:Gene: F13011.10  
A:Map position: 1

Query Match 1.9%; Score 169; DB 2; Length 2698;

Best Local Similarity 18.8%; Pred. No. 0.023;

Matches 385; Conservative 269; Mismatches 674; Indels 720; Gaps 97;

Qy 158 YREVADCVDKLEKELSVKROQFEELYKTEAPMET--HGNLMTERQVSRWFVQGLRQS 216  
Db 844 FKMLQDLPRLSHDML-----ETDIKIPHTREGMLLSRGV--YVAQITGAK-- 890  
Qy 217 MLEETIPIYAYFEMAPBDLVLTQMFKEQSGSGQT---NRHLVDETMPFVDRIQYF 272  
Db 891 -----YTKQEPSSNSHLKK-----GLASSETANSGRDPAKLTAKKADKGTAK 934  
Qy 273 SA---LILVEGMDISLHKCALDRRELHOPADGL-----ICQDMDCLMTFGDIPHPAP 325  
Db 935 EARELMLEBASTRNVRIQKSLSLVHALGEMGLAPVCHSGLPLPLATFLDPLRSF 994  
Qy 326 VLLAWA-----LRLHTLNP-----ETSSVVRKIGGTALQNLNVQY-----LTR 364  
Db 995 IVSAAAFENLVLAARCTQPLCNMMLAISTALRLIAIDVDTLS--PDFRPSYDKAGKTYEG 1053

Qy 365 LLQSLASGNDCTTSTACMCVYGLL-----SFVLTSLEHLTGNOQDIIDTACEVLADPS 419  
Db 1054 LPERIVNG-----LSISCSGSLPYDPTFFIFPVL--YHVLG-----VVARQASVQPA 1100  
Qy 420 LPELFWGTETPTSGIILIDSVCGMFPHLLSPQLLRAL-----VSGKSTAKKYVSFLDK 474  
Db 1101 LNELCLGLQAD-----VANALYGVYSKDVHRLACLNNVKCI PAVSKCSLPQNV-----K 1151  
Qy 475 MSFNELYKHKPHVDVISHEDDTLMRQPTKLLYPLGG-----QTNLRIPQGTGVGM 526  
Db 1152 IATNIMIALHDPEKSVASADLMARYGHDIGTDYSGIFKALSHININLRAAABA----- 1207  
Qy 527 LDDRAVLYRWMEYSVSWTLFTCEIE-----MLHVSADVDYQ----- 564  
Db 1208 LADALHESPSSIQSLSTLSTFLSYTRDATSGEDVDAGMIGROGIALALQSAADVLTITDL 1267  
Qy 565 -----HCRVXPIIDLVHKVISTDLISAD-----CLPLI----- 593  
Db 1268 PAVMTPLISRALADPNTDVRGKMINAGIMIIDKHGKENVSLPLIFENYLKKEASDEBEY 1327  
Qy 594 -----TSRIYMLQRLTYISFPVIVASCVN--CLT----- 623  
Db 1328 DLVRGVVIFPTGALAKHLARDPPKVHNVKLELVLTNPSSVQRAVSTCLSPVLISKOL 1387  
Qy 624 -----VAAANPAKVMTDLRHTG--- 641  
Db 1388 MKSDKYGERRGAAGLAGVWNGFGISLKKYGLVTLQALIDRNSAK-----REGALL 1442  
Qy 642 -----PLPFVAHPVS-----SLSQWISAEGNNAAGNYGLNMSSEOPQGEYVTI 685  
Db 1443 AFECICEKGLGLFPFPYVYKMLPLLVSPDQGVAREAAFCARAMMSQ---LSAYGVKL 1499  
Qy 686 AFLKLLITLVYG-----QLGSTQSG-----LVPCVMVLKEMLPSTYK 725  
Db 1500 -----VLPSLLGLBDAKWRRTQSSVQLLGANAFCAPOQLSCLPRVVKPLEVLTDTTP- 1554  
Qy 726 RYNSHG--VREQIG-----CLILELI-----HAILNLCHETDLSSHTPS 763  
Db 1555 KYQASAGLALQOVSVIKNPETISSLVPTLLALDNPETHTALDITLQTFVNSVDPS 1614  
Qy 764 LQPLC-----ICSLAYTE-----AGQTVINIMIGVDTI 792  
Db 1615 LALLVPIVHRGLRERSSETKKKAQIVGNMCSLV--TEPKDMPYIGLLPVPKVLVDPI 1673  
Qy 793 DMVMAAPRSDGARGQGGQ-----LKITKYKLAFSYNNVYIRKPSNVVSPLEQA 844  
Db 1674 PEVRSVAPARAAGSLIRGEGEDNFPDLVPMLEPTELK---SDTSNERY---GAAQGLSEV 1726  
Qy 845 LSGHGA--HGNMLIVLAKYIYHKHDP-----LPR--LAIQLKRLATVAP----- 887  
Db 1727 IALGTDTYFENILPDLIRHCSHQASVVDGYLTLPKFLPRLGAGQFOKYLQDLVLRPALLDG 1786  
Qy 888 -----MSVYACIG--NDAAIRDA-----FLTRL 909  
Db 1787 EPIFGRLTSELFPDSITKLVLVYDVLGADENESVADALGAGHVLVEHHAATSLPL 1846  
Qy 910 QSKIED-----MKIKWILEFL-----TVAVEQPGILEFLNLEVQSGD--GSKFSS 956  
Db 1847 LPAVEDGIFNDNMWIRIROSSVELLDLLEKVAIGTGKALLE-----GSGDDREASTEA 1898  
Qy 957 LGMMSCUAVVELIDSQOODRYWCPLLHRAAIAFLHLMQDRDSAMLVLRTPKPFEN 1016  
Db 1899 QG-----RAITDILGMDRNE-----VLAALVNRVTDVSLSVQALAHVKT 1940  
Qy 1017 -----LTSPLFGLT--SPSESTSEPSIETCALIMK-----IICLEIYVVK 1056  
Db 1941 IVANTPKTLKEIMPLIMSTLSSLASPSESRQYAGSGLVYKLGKRVLPFLIPILSK 2000  
Qy 1057 GSLDQ-----SLKDTLKFSIEKRPVYMSGYKSL-----AVHVAETGSSCT 1099  
Db 2001 GLKDPDVDKROGVGIGINEVVASAGRQLSPMQLIPTITRALCDSLLEVRBSAGLAFS 2060  
Qy 1100 SLLEYQMLVASMRMLLIATTHADIMHLTDSVVRQFLVDVLDGTKALL-----L 1149

Db 2061 TLTKSAGLQA---MDIIIFP-----LLEALEDDMSSTTALDGLKQIITSVPTAAVLPHI 2110  
Qy 1150 VP-----ASVNCRLRSGMK-----CTLLILLRQMKRELGSVDEILGP----- 1187  
Db 2111 LPLCLVHLPLSALNAHMLGALAEVAGAFNTHLITGLTILPALLSMGGENKEVQELAQEAAR 2170  
Qy 1188 -----LPEILIEGVLOAQOOLMEKTKAVPFAFITYLQMKEMK---VSDIPQY 1231  
Db 2171 VVLVLDSEGEVETLISLLSLKGVSDSQASIRRS-----SAYLLGYPFKSKLYLIDEPAP- 2222  
Qy 1232 SOLVLNVCEITLOEVALFDQTRHSLA-----LGSATED-----KDSMEPDD 1273  
Db 2223 -----NMISTL---IVMLSDSDSTTVAWSMEALAVISVPEKVLPSYIKLVRAVST- 2272  
Qy 1274 CGRSRHRDRDQG---VCVGLGHLAKELCEVEDGDGSMLOVTRRLPLPTLLTTL---EVSLL 1328  
Db 2273 -ARDKERRRRKGGYVYIPGLCLPKSL-----KPLLPVFLQGLISGSABL 2315  
Qy 1329 RAKONLHFEATYHLTLTLARQOGATVAGAGITQSTCLPULSVYQULSTNGTQOTSAS 1388  
Db 2316 R-----EQALIGIGELIEVTSQALKEFVYIPITG---PLIKRI--IGDRFPVQVRSAT 2362  
Qy 1389 RKSLLAPSPWPGVYRLSWSLMEQLTKLRYNFLPEALDFGVHQBRTLOCL-NAVITYQSL 1447  
Db 2363 LATLLTLL-----IQRGGMALKP-----FLPQL-----QTTVPKCIQDSRTIRISS 2402  
Qy 1448 AGLSEADHTVGPILLOLSNPKMEWHHLPLQMRDIOVNLGYLCOXCTS-----LHSHRRM 1501  
Db 2403 AAV-----ALGGLSALST-----RIDPLVGDIMTSP-----QADAGVEBALISANRGV 2446  
Qy 1502 LOHYLONKNGDGLPSAVAQVORVPPSPASAASSSKQPADPTEASBOQALHTVQYGLTKI 1561  
Db 2447 IKH-----AGKSIGPAVAVRI-----FDLLKDLMHNEDDQVRIASITSMVLGV 2487  
Qy 1562 LSKTLAALRHPTPDVCQILLDQSLDL-----AEYNFLPALSTPTPTFDEVAAPS--FG 1612  
Db 2488 LSOYLEAAQO-----LSVTLQEBVNDLSAQNMWGARRHSVCLISILLKHNPSITMTSLSFS 2541  
Qy 1613 TLLATVAVNALMNLGELDDKKKEPLQAVGLSTQAGSTRITISLMTFMENCFYLLISQAR 1672  
Db 2542 SMANSLK-----SLKDKKFPLEBS--STVALGRLLLKQL----- 2574  
Qy 1673 YLRDPVAVHPDKQRMKQELSSSELSTLSSLSRYFRGAPSPATGVLPSPQKSTSLSKA 1732  
Db 2575 -----ATDPSTKTVIVDIVLSIVSALHDSDSE-VRRRLSSLSKAFAMDNSATMANISVI 2628  
Qy 1733 SPESQEPRL 1740  
Db 2629 GPPLAECL 2636

RESULT 11  
T29042  
hypothetical protein B0228.3 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence revision 15-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T29042  
R:Leimbach, D.  
submitted to the EMBL Data Library, March 1995  
A:Description: The sequence of C. elegans cosmid B0228.  
A:Reference number: Z18324  
A:Accession: T29042  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-4365 <LEI>  
A:Cross-references: UNIPROT:Q81G62; EMBL:U23166; PIDD:MAC38807.1; CESP:B0228.3  
C:Genetic: 3  
A:Gene: CESP:B0228.3  
A:Initrns: 2545/3; 3781/3

```

Best Local Similarity 18.5%: Pred.No. 0.069:
Matches 368; Conservative 296; Mismatches 646; Indels 674; Gaps 95

198 LATEROVSWPVQICR-----EGSMLEIIFLYYAIFEM--ABSD 235
      : : : : : : : : : : : : : : : : : : : : : :
169 IRRERSEKVFQO--LRASADEITIREVALGSMWOLEQAFAFMILITTSRCDLRTIASN 227
      : : : : : : : : : : : : : : : : : : : : : :
236 LVLVTMF-----KE-----QGFGRQNRHLVDETM 262
      : : : : : : : : : : : : : : : : : : : : : :
228 IIANTEIFPDVAEEKMSVGMWKREKREKHSKFTSREELIQFWMKGEDEERVVKTL 287
      : : : : : : : : : : : : : : : : : : : : : :
263 DPFVDRIGVFSALLIVEG---MDIESLHKCALDRRE----- 296
      : : : : : : : : : : : : : : : : : : : : : :
288 K---PRMVSFKSLSEVAASSTSENVSLGMWKADQKSEFTISQKLSPEKIVSEAYGSES 344
      : : : : : : : : : : : : : : : : : : : : : :
297 -LHQAQOGLICODMCLMTPGDI PHHAPVLAMALRHTLANDE--TSSVVRKIG-- 350
      : : : : : : : : : : : : : : : : : : : : : :
345 KLDQFFQ---VWEKMDWSNIEIGEENHA---LSANLR-FLADEQKVCDDIGLCKLAPK 396
      : : : : : : : : : : : : : : : : : : : : : :
351 -----GTAIQ-----LNVFQYLFRLLQSLASGQN-DCTTSACMVCYGLSPVLTSLEL 398
      : : : : : : : : : : : : : : : : : : : : : :
397 QODESVGTQIQEIRATAVMVSVRASLLITTSINSFSGNISDSKAVFNSLIVTISHNL 456
      : : : : : : : : : : : : : : : : : : : : : :
399 HTLGNQODIITDCAVGLADPELPFLFNGTEPTEGLGIILDSVCGMFHLLSPILQLRLAL 458
      : : : : : : : : : : : : : : : : : : : : : :
457 STIGSS---ETTASISYODIPEWL-----AASKLWISRNSE-FLKKEIREPVIQVESF 508
      : : : : : : : : : : : : : : : : : : : : : :
459 VSGKSTAKKVSFLDK-----MSFYNELYKHKPHDIVSHEGGLMRGQPLVPLDGQT 513
      : : : : : : : : : : : : : : : : : : : : : :
509 WNTINDQKVAVLNKEKDIYSYSLNLTILASMEHEIISQF---LVNHTEN---AGKV 559
      : : : : : : : : : : : : : : : : : : : : : :
514 NLR--IPQGVGVQMLDRAYLVWEXSVYSWTLLFTCEIEMLLHVVSADVDVGHQCRVKP 571
      : : : : : : : : : : : : : : : : : : : : : :
560 NIKRISPSEILVS-----SSFKITSSDLOQFNVLEKMDMSQ----- 595
      : : : : : : : : : : : : : : : : : : : : : :
572 IID---VHKYISTDLSDIADCLLPITSRIYMLQRLTTVISPPVDVASCNCITVLAAAR 628
      : : : : : : : : : : : : : : : : : : : : : :
596 -ISLPAQEHALISKVR--SLASSAEL-----SNLIGKLRAP 630
      : : : : : : : : : : : : : : : : : : : : : :
629 NPAKWTDLRHTGFLPVHAPVSSLSOMISAEGMNAGYNLMMNSQC--POGEYGV 684
      : : : : : : : : : : : : : : : : : : : : : :
631 EMQETIEDLKRQAOQ--ARVVLNVSSMS-----NTISSSEFRIENKATL 678
      : : : : : : : : : : : : : : : : : : : : : :
665 IAFRLITTLVKQGLSGTSO-----SQGLVPCVWFVK-----EM 718
      : : : : : : : : : : : : : : : : : : : : : :
679 TNLGIIITTDPLSTLGASSNQITENDYTEAKELEASKVLVKNLSVLTKEIRRESGDEV 738
      : : : : : : : : : : : : : : : : : : : : : :
719 LPSYHKRRVNSHGVRGIGCLILELHAILNLCG-----END-----LHSH 760
      : : : : : : : : : : : : : : : : : : : : : :
739 VQGF-----WNTASKEKVGAILCEKLSIHHTLHAIKRVTSLSLTDIOEKQOKLSTHH 794
      : : : : : : : : : : : : : : : : : : : : : :
761 TPESLQPLCISLAYTEAGQIVINIMGI-----GVDTIDMV-----MAAPRSDGAE 806
      : : : : : : : : : : : : : : : : : : : : : :
795 FVKLSTRDVVOAFAFRISSESVDHLTLVEKMDKSIDLVEPVYSNISSNVQALATSSSQS 854
      : : : : : : : : : : : : : : : : : : : : : :
807 GQGGQGLL-----IKYVLAFSV--TNNVIRAKPSNVVSLPEQLSLQH 848
      : : : : : : : : : : : : : : : : : : : : : :
855 SGILGKLPPAPEBETTSKEIRINTAKCLIANVTSSLSNLTSSDASLDMISEQOAF-- 912
      : : : : : : : : : : : : : : : : : : : : : :
849 GAHGNNLIJVLAKYIYHKHDPALPRLAIQLLKLATVAPMSVYACIG-NDAAAIRDA-- 904
      : : : : : : : : : : : : : : : : : : : : : :
913 ----SNIIIGWA-----SNMLTSSQSSTSSSGFNNVVELSARV 949
      : : : : : : : : : : : : : : : : : : : : : :
905 -----FLTRQSKIEDM-----RIKIMILEFL-----TYAVE--TOP 934
      : : : : : : : : : : : : : : : : : : : : : :
950 LQBSNKOMLQKQRESSEELIHGIWSTASETEVALVVEKLEYHQAOKTLAIQWATOS 1009
      : : : : : : : : : : : : : : : : : : : : : :
935 GLIETPLNTEVKQSDGS---KEPSIGMWSCLHA-----VLEIDISQOQDRW 979
      : : : : : : : : : : : : : : : : : : : : : :
1010 VNSELL-----GSGGNLAKISTLPLPRVIVISAAFGISNEVIOKMLEVLTIKGEXSTIS 1062
      : : : : : : : : : : : : : : : : : : : : : :
980 CPELLRAAIAFLHLMQDRDSAMLVLTPEKPEMENTLSPLEGLTS---PSEETSEBI 1036
      : : : : : : : : : : : : : : : : : : : : : :

```

D	b	1065	LPQSEYBISQNVRL-----ABSF--NCDS--ILGKINAPGQTEHTSHL	11069
Q	y	1037	LE--TCALIMKICIEIYVVGSLDGLKDTLKKFSIEKRPAYSGVYKSLAVHAETE	10944
D	b	1107	IEKTPAAVAN-----YKAIVESVIS--KSSIAKLPADEKAITITLAGVLISKD	11544
Q	y	1095	GSS--CTS-----LLE-----YQMLVSAMFML	11144
D	b	1155	VSSLCSTSETEGFPDQRIFRNONANISLGAVSEITLLQRLREPIENQVOGFWSTASNOEKA	12144
Q	y	1115	LIATTHADIMHLPDSVVVRQQLFDLVDGTALALLVPASVNCILRGSKCTILLILRQW	11744
D	b	1215	SFIKKQKIDMYDMKVAIAIQMISTIDGD----PSSASISEFTFNKIDKAREV----	12666
Q	y	1175	KREIGSVEILGCLPTEILEGVLAQDOQMEKTKAVFSATITVQMKEMVSDI-----	12288
D	b	1267	AAEFGIANESVOKALLEVANKVEMSDISLPEKLGHQNI--SSNVRVLADSSYNCDSIIGKUNA	13255
Q	y	1229	--PQYSQVLVAVCEPFLQREVIALPQGTNHLALSGATBDKXDSMETDCCSRHRHDQRGV	12866
D	b	1326	PEPQSAVDQOISEHQTELEVIANIKSAVSV-----ISNDVLSKRSBEKAVI	13744
Q	y	1287	C-VGLHLAKELCEVDEGDSW-----LQY--TRRL--PI-----	13166
D	b	1375	SNINLIIISCCLASMSSTNSPEPQRILEPQNNAVILGTPASQVFTNLDQFIQSEVQ	14344
Q	y	1317	-LPLUTLTLEVSLRKNQNLHF-----TE--ATLHLITLARTQOGATAVAGACIT	13635
D	b	1435	FWSTNSQEKASLVIRQKLANKVYDAKMMKATEIAVQSLNSINTSEHTLIESQKSPKSI	14944
Q	y	1364	QSI-----C-PLSVQVQST-----NCTAQTPASRSLSLAPSGVYR	14024
D	b	1495	ESIMASFGVSEBHOKTLEILSKSELSTIVLPAVEYKATVQ---NIALSEPNPN----	15466
Q	y	1403	LSMSIMEOLLTLRYNPLREALDFVGVIQOERTLOCLNAVTV-----OSLACLEBAD	14544
D	b	1547	-----CESITLCKLANPDPQSAVMQDIIMNEQKRLTVVSNIGSLAESVISNDSLEKLETV	16014
Q	y	1455	HTVGFILOLSNPKMEWHPLPOLMRDIOVNTGYLCOACT--SLL--HSRKMLOHYLONK	15095
D	b	1602	KSV--ILKISERI-----VSHDITLALSSCSDPNLLQESNTQEKADIFLXTP	16466
Q	y	1510	NGDGLPSAVAQRYQRPSSA--SAAPSSSKOPRADTEASEQALH-----TVQ	15555
D	b	1647	NSQMLIERLREPIEQVOGFWSTASNSKEQMFLEKYE--TIHAMLOTFBSASLVSETVQ	17044
Q	y	1556	YGLKILSKTALAALHFTPDVCOIL-----LDPS--LDLAEY-----N	15931
D	b	1705	RDMFATV-QSLAALRSIQLTPREILCAFGISNBEIQTFPGDLBVDVMSHIDIPICALBDN	17666
Q	y	1592	FLFALSFTTPFDSEVAPSGFTLLA-----TVNALMMLGELDKKKEPL	16355
D	b	1764	LLVNLRIV-----MADAPNVFGLISRPDENQETSTLEKRNRIQPLNMLQRLDQTVQ--M	18148
Q	y	1636	TQAVGLSTQAGSTRILKSLMLFTMNCVYLIIISOMRMLRDPAYH-----PRD---	16835
D	b	1819	TSSLTRSDERIEVKVSNIIISLSSSN--LGDLLSQVQ--LAQPNMSDEFKSFELPRNILL	18766
Q	y	1684	-----KQRMKQELSELSTLLSSLSISRYFRGAPSSPATGVLPSPQOKS	17266
D	b	1877	NRVVPETSEBSIQSPWKTQSQALEBASSTISEKLSMLQSEF-----VSSAAKQVS	19255
Q	y	1727	TSLS	1730
D	b	1926	TSLT	1929

RESULT 12  
T28915  
hypocherical protein C13F10.4 - *Caenorhabditis elegans*  
CISpecies: *Caenorhabditis elegans*  
CDate: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #next\_change 09-Jul-2004  
Accession: T28915

R.Tin, A.; Wohldmann, P.  
submitted to the EMBL Data Library, April 1997  
A.Description: The sequence of C. elegans cosmid C13F10.  
A.Reference number: Z20543  
A.Accession: T28915  
A.Status: preliminary; translated from GB/EMBL/DBJ  
A.Molecule type: DNA  
A.Residues: 1-2076 <Tin>  
A.Cross-references: UNIPROT:O01485, EMBL:U07006; PIDN:AA647965.1; GSPDB:GN00023; CESP:CL  
A.Experimental source: strain Bristol N2; clone C13F10  
C.Genetics:  
A.Gene: CESP:C13F10.4  
A.Map position: 5  
A.Introms: 114/5; 264/1; 319/3; 358/2; 664/2; 756/3; 840/3; 1022/3; 1119/3; 1840/1; 2056

[illegible]

```

Db 696 -----HKSFEQSYAAL---LRELVADITLSDNSQSTLTLSLPISGFTGVCKILISPYDA 747
Qy 1229 PYSQYLVANVETIOEYIALFDQTRHSLAGSTEDKSMETDDCSRRRDQDGV 1288
Db 748 TDYSNV-----EDLIQTPIS-----SVSLGNIEEDLSNLIIRASAS-----782
Qy 1289 LGLHLAKELCEVDDGDGSMLO-VTRRLPILPTLLTTELEVSRLMKONLHTEATLHLLTL 1347
Db 783 -----QIGTWPENDEPFLCTLNTALTLYKVPFLVNNKHKVOITHEPMDTI 829
Qy 1348 ARTQOQATVAVAGITOSICPLLSVYOLS-----TNGTAQTPSARSKSLDAP 1395
Db 830 QSKS-----NAGRKQAILVNALTAKLISYKLCQGRGHKLDMETLQ-----RASFDL- 876
Qy 1396 SWPQVRLSMLEOLKTLKALNPLPRLDVGCHQETLOCIAVAVYOSLA-----1448
Db 877 -----ISSLSNSCPTWRLVG--AEAL-----ARLSQAVNSPPFVASIAQYCDKL 920
Qy 1449 ---CLEADHTVGFILQISNFKEMWPHLPOLMRDIOVNLGYLCOACTSLHSRKLQHYL 1506
Db 921 NSCQDEINRS-GHYLALGCL---HRHVGSL-----946
Qy 1507 QNKNQDGLPSAVAGRVORPPSASAPSSKOPADTEASEQOAL-HTVOYGLKILSKT 1565
Db 947 ---GSG-----QHLNTVGSVVLALAESKMPKQVCALVAMALIARETSGMFRFVET 996
Qy 1566 LAALRHPFDPVQCILLDSDLAELNPLFALSFTPTPTDSEKVASFG---TLATVAVVA 1621
Db 997 -----TLSSCLKL-----ISPTFVAVVOGISKCLTALTTCVQPE 1033
Qy 1622 LMLGELDKKEPTTOAVGLS-----TOAEGFTLKSILMTFMCNCFYLLISQAMRYL 1674
Db 1034 LSCGVIDGVATSLLAACAQIQLSHSDPVOQAEATISGLQOMLF-----APRY- 1080
Qy 1675 RDPVHPDKQRMKQKSELSTLSSISRYFRGAPS-----SPATGVLP 1720
Db 1081 ---VH-----MAQLVVDISSLSSTHVLVIRROVSCLRQLVORSEKVRNHAQVLP 1129
Qy 1721 SPQKSTSLSKAPES-----QEBPLIOYQ 1745
Db 1130 QGIYDVTNKKKRALPESGLEGALFGMLDTEVNEKEIRCHLQETLISLQ 1176

RESULT 13
AS7036
talin - glime mold (Dictyostelium discoideum)
C/Species: Dictyostelium discoideum
C/Date: 03-Oct-1995 #sequence_revision 03-Oct-1995 #text_change 29-Oct-1999
C/Accession: AS7036
R/Kreitelmeier, M.; Gerisch, G.; Heizer, C.; Mueller-Taubenberger, A.
J. Cell Biol. 129, 179-188, 1995
A/Title: A talin homologue of Dictyostelium rapidly assembles at the leading edge of cel
A/Reference number: AS7036; MUID:95213284; PMID:7698984
A/Accession: AS7036
A/Status: Preliminary
A/Molecule type: DNA
A/Residues: 1-2491 <KRE>
A/Cross-references: GB:U14576; NID:G797298; PIDN:AA46586.1; PID:G797299
C/Genetics:
A/Introns: 131/1; 222/2

Query Match 1.7%; Score 155; DB 2; Length 2491;
Best Local Similarity 17.9%; Pred. No. 0.19;
Matches 276; Conservative 227; Mismatches 501; Indels 542; Gaps 72;

Qy 560 ADVTQCRVPRPIIDLVHAKVISTDISIAD-----L 590
Db 408 ANQOPQSQI-PIITLKSALRATDILLIGELGFRSGTATPQNFTRSTTLTPQOFKQL 466
Qy 591 LPIIRITM-----LIQRLTYIAPVVDVYASCVNGLT-----VLAARPAKVMYDILRTG 641
Db 467 ISHTNAMAIAAQGLFQDMWT--PPPTGGIAAFQQAIRRAQIIMAEIN--TVGTAAKNAG 522

```

```

Qy 642 FLPEVA---HPVSLSQMISABGNMAGYNLMMNS---QPOGEYGVITAFILITLLV 695
Db 523 YFPDMSFSPBEITIGVAFKLSESMARLLAISTIGTCTDCKSKQAAATELFNNQSLTLM 582
Qy 696 KG-----QIGSTOSQGIIPC-----VMFLKEMLPYHKRMYNSGV 732
Db 583 MAACDNEYVDSSSKLIECAKXVSAIAADMLVYGNKSEFIDELLGQIQNTLKSTSL 642
Qy 733 REQIGCILLIHALIINL-----CHE-----JDLSSHPSLQFL 767
Db 643 TSD-----BLSTTEMASTSCPSERKQITNTIGALNOSNALLTAFPSGEIPEQY- 695
Qy 768 CIGSLATYAGQTV--ININGIGVDITDMAAQPREDGAGGQGGOLLTKT-----VK 819
Db 696 ---NLNARSDIIESVNLINLVAMDGEREYKISITNGVE-VGEGGILAGTNLTSEFAT 751
Qy 820 LAFSVTNVIRLKEPNSVSP--LEQALSOQHAGNNLAVLAKYIYHKNDPALRLAIQ 877
Db 752 VANDLTNAIMTMR--SNLKMPDTVMESYKQVAGHANRLI-TCTKAVASRAD-----TQ 801
Qy 878 LKRL--ATVAPMSVYACIGNDAAAIADAFILRLQSKIEMRIKVM--LEPLVAVETQ 933
Db 802 SQQLFNSTNAVPEESVANLSNHC--SYIKPEQEAHTFOIETAGHLOFTLQNNSTD 857
Qy 934 PGLIEFLNLEVKDGSDEKFSIGMWSCLHAYLELIDSOQODRYMCP-----981
Db 858 AGRIACTISLR-----DYSKEM-----IAQVSSLSTRTSQVLPDANGITLLGAK 905
Qy 982 ---PLHRAAIAFLHLMODRDSAMLVLRTEKFEW---NLTP--LEFTLSPSET 1031
Db 906 DVSPALSKLWVGIRKVLDPKSEATQWELLTQKOSLPPMNLVSTCKRPAPKISDPNQK 965
Qy 1032 ---SEPSILETCALIMKI-----IC-----LEIY-----YVVKSLD 1060
Db 966 QRLTFSSDAAQSQVKLMKAGAYKICGHIETEBALFVDPSTIADETTEIATAGFLD 1025
Qy 1061 Q-----SLKDTLKFSIEKRFAYWGYKSLAVHVAETEGSSCTSLLEYQ 1105
Db 1026 AVSGTREGAELLMVAIKD-LNKVNE-----LVTDIRVAPRL-GDLVGSATESA 1075
Qy 1106 MLVSAMKMLIIRT--HADIMHLTDSVY-----RRLFL-- 1138
Db 1076 SVAISAKTILICATGQOVQTKMGITKQIMDEQILIRASRSVSNPNDRSLELDRR 1135
Qy 1139 --DVLDGKALLVPAVSNCLRLGSMKCTLLILLRQMKRELGSVDLEGLTE-----1190
Db 1136 SNDVISITALLVGSITANVDCKELDEASADISNL--SLKMSLESISQPTSEFAYV 1191
Qy 1191 ---ILEGVLDQDQIMERTKAKVFSAFITVLQMKEMRVSDIPQSQ--LVINVCTELQ 1243
Db 1192 EIASSTKALNAAQVAVAMARNKRLKGLASAKITASALSTVSHAQNAIVLTENEATK 1251
Qy 1244 EGVALDQDRHSLAGSATBEDDSMETDDCSRRHDDQDGVCLG-----HLA 1294
Db 1252 NALIA-----STVALDG-----QIGLDFSKRIANYDPIYDQNLINQAKVEDHLV 1300
Qy 1295 K-----ECVEVDGDSWLQVTRRL--PILP-----LTLTELSLRMKONLHTE 1338
Db 1301 KVGSLGDDGNRTIC--DEAVDRITTEATRSIDKTLITLDTSGLOINAHLEMHQOSL-----1354
Qy 1339 ATHLILLTLARTQGA-----TAVAGITQS-----1365
Db 1355 ---LATQASKKLGISITSNLVNSKNNSDVGSSTDAERLIIEIAKHVHCSISTYN 1410
Qy 1366 ---TCLPLLSYQSLSTNGTAQ-----1383
Db 1411 PDILIPAKSILDSQMLTANQADVNVLSHAATTAQTOQLGITRRASQFNEQEQOV 1470
Qy 1384 -----TPSASR--KSLDAPSPGVYRLSNLSIMEQLKTLR-----Y 1417
Db 1471 QVRDGIYKSTOQLAHATISSARAVKSVTSKE-PG-----AKAMISQSLKODESALNNLLIT 1525
Qy 1418 NFPEALDFGVHQ-ERTLQCLNAVRT-----VQSLACLEADHTVGFILQISNFKEM 1470

```



```

Db      1526 SSVPSRSGIGIADPNKLMSTCRSVSTASQULIISASSCSQKPD-----IELSLISEN 1580
Qy      1471 HFLPQLMRDI-----OYNLYLCOACTSLIHSKRM 1501
Db      1581 AVLMTNSIKDIIKVTSSMMPGVNCFCEAIEIAQRAISDLSSVALSAVAGSPSSANKEG 1640
Qy      1502 LQHY-----LQNKDGLPSAVAQVORPPS---AASA-----AP8--SSKOPADT- 1543
Db      1641 LSHVESQRLVDVTKITGTINDLKASRQSPBAIGISAKALSPFAPLVVTKPALATA 1700
Qy      1544 -EASEQCALHT-----VOYGLKILSKTLAALRHFTPDVQCIILDDSLDAEYNELFALSF 1598
Db      1701 PDAAQNDLVTESKNVGSILKLCQASLIASSNPKETQIIVKCVADSE----- 1751
Qy      1599 TTPFPDSEVAPSFGTLATVNAVANMLGELDKKKEPLTQAVGLSTQAEQTTLSLMFT 1658
Db      1752 -----AMSLVAVQISSGVNLKXELDESILDRIRKSV-VQTSAKDA----- 1789
Qy      1659 MENCFFYLLISQAMRYLRDPAVHPRDKQ-RMKQELSESTLSSIS 1703
Db      1790 -----PKQSENRYQEQYKEBELSLNLTKNLA 1813

```

## RESULT 14

```

T37964
probable ubiquitin ligase - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C/Accession: T37964
R:Devlin, K.; Churcher, C.M.; Wood, V.; Bartrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, September 1997
A:Reference number: Z21757
A/Accession: T37964
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-3227 <DEV>
A/Cross-references: UNIPROT:O13834; EMBL:Z99531; PIRN:CAB16714.1; GSPDB:GN00066; SPDB:SF
C:Genetic:
A:Experimental source: strain 972h-; cosmid c19D5
A:Gene: SPDB:SPAC19D5.04
A/Map position: 1

```

```

Query Match      1.7%; Score 154.5; DB 2; Length 3227;
Beet local similarity 18.3%; Pred. No. 0.31;
Matches 347; Conservative 255; Mismatches 665; Indels 631; Gaps 89;

Qy      54 YKPPSPSSAEKVK--ANKDVASPLKELGIRISKFLGLDEQSVQULQCYLQEDYKGTGDS 111
Db      12 YSQPPRAVAFIRQAQNEVTS---DLGL-----VSLCSFRRKNDMPYPRGD 55
Qy      112 VKTVALQDERGQALIKTADYY-----YEERTG-----ILRCYALHLLTYFQDERH 156
Db      56 LYSWVPVLRPDAIERIVHEHSLKDKVQTKPFDSDTLILLEILISFAHLHSKAN-RS 114
Qy      157 PYR---VEY-----ACVCK-----LEKELVSKYRQOF-----E 182
Db      115 IYNSTVYLEYLINSSVLEVIDSTLALLHLIVQKATISRRGQSLSODRLFRFLMFLPQ 174
Qy      183 ELYKTE-APTWEHTGNLMTERTOV-SRWVQCLREOSMLLEIIFLYAYAFEMAPSDLVLT 240
Db      175 DAMKTGFSQVNET---LFSNEIPEW---CSLEIS-----YKSSPSKDFESA 217
Qy      241 KMFKEQFGSGSRQTNHNLVDETMDFVDRIGYFSALILVEGMDISLHAKCALDRRELHQF 300
Db      218 SQNPSEGS-----ILKLPYKVKL---GKPIBELLVKTLHDNQIPRQY 257
Qy      301 AADGLICQDMCMLTFBDIPHHAVVLLAMALLRITLNPETSSVVRKIGSTALQANVQ 360
Db      258 SFD-----LLVSLMLRQNLVDINRRRLMIRIGLALASNLVYA 294
Qy      361 Y-----LTRLQSLASGNDCTTSTACMCV----- 385

```

```

Db      295 HSOAVQTRFLIADPEITHTLNLVSPDVVLPQNFKAVCFECFKAFFPKSMIPSVLASLN 354
Qy      386 -----YGLSLFVL-----TSLEHTIGNOODIIDTACEVL-----ADPSLPBLFWGTEPTG 433
Db      355 VSVSYGLMMNLVRFDSKLENPNFYEREYDSEYDFLQFMFTSSPLGGNMACSAGLSL 414
Qy      434 GIILDSVCGMPHLLSPLOLLRALVSGKSTY-----KKVYSGFLDKMSFYNEL----- 481
Db      415 GYHLSVTKPQATYVVAISIVLHDHIDYSNAFAFDPSBSKDLMDLVDRVQYLEAGLQDI 474
Qy      482 -YKRPADV-----ISHEDGTLMRQTPRLVLP-----GGQTNLRIPQGTVGWMLDDR 530
Db      475 KSGKGNPEIIVANMDYALSYDRYFLKNLKFVLIHLOGSGSVEIR---NLIDSLISL 531
Qy      531 AYLVRMEISYSSTWLFCEIML---LHVSTAVIYQCORVK-PIIDLHKV----- 580
Db      532 AFLLEHHEVVS-NLFASSTYNIMSTFIHNEPTCGIIEKRLSHAFLDAVNRKLINSDA 590
Qy      581 STDLSIADCLPIYSRIY-MLQRLTVYISPPVVIASCVNCLTVLAARNPAKWTD--- 636
Db      591 ITSIPFAGALCINSQGFPLEK--NPQPQSEIFTSLNCKSLISSDPAIIGTYIDE 648
Qy      637 -LRHTGFLPFVAHPVSSL-----SQMISA--EGMNAAGYGNLNNSEOPQEGYVTL-AFL 688
Db      649 LMRHQ---PSLKQDPIVKMIFKACQVGSALDNFNPQIYN---KBYVYLLYLETFSSFL 702
Qy      689 RLITLVKQGLSGTOSQGLVPCVNFVKEMLPY--HKWRYNSHGVBQIGCLILEIH 745
Db      703 ENIIT-NEGHAARYLISKGIYSHVNLIOHPLAAGFIDSSAFNSFFV-----LLH 751
Qy      746 AILMLCHETDHSHTSLOFLCISLAYTAGQTVINIMIGVDTIDMWAAQPRSDGA 805
Db      752 -----HAYFDABEV--FRPLDCLITRCEGTEFTFIVSLKQNTIS----- 791
Qy      806 EGQGGOLLIKTVLAASVTNNVRLKPPSNVVSLEPQALSQGA---HGNMLIAYLAKY 862
Db      792 -----LIDSMNGHEDANNPLHFSIVGNLITFPALEFSHAALKKAGN--LPVLQLF 841
Qy      863 IYKHGDRPLPRLAIQLKRLATVA-PMSVYACLG--NDAAIRDAFLRLQSKIEDMRK 919
Db      842 I---SPSRVAGIPDIICNIKSIATSLDIHICLGVSDPVLCSLSLTITIVDKQEKKEFE 897
Qy      920 VMILE-----FLTVAETQGLLEFLNLEVKQSDGSKESFGMNSCL 963
Db      898 TKKELIYODSFCFPQINRSNFSQIAY---GVSKFTSLTRALGNISVQDFN--EYMI 951
Qy      964 H-----AVLELIDSGQODRWYCPPLHRAAIAFLHALWQDRDSAMLVLRTPKPE 1015
Db      952 HKGSLNALVVDDELINSSKQITSHPSL---SIASLEA-----SLIFVLGASSIRE 1001
Qy      1016 N-----LTSPLFG-----TIS-----PSETSEPSILETCALIMKI 1046
Db      1002 DSKVTLVLLISRLGGCRTDVLISINETYSGFRLSDRDLPSKSNKVLALSTSLNLT 1061
Qy      1047 ICL-----EIVYVAGSLDOSLKDILKKFSIEKRAVNSGYKSLAVHALETGSSCSIL 1102
Db      1062 ILVFTSADPNSEISKTLAMALKSEFDMTF-----NNSGSKLHVL 1102
Qy      1103 EYQMLVSA--WRMLIATTHADIMHLDTSVVRRLFDVL-----DGTKALLVPA 1152
Db      1103 HARIFISVLMR-----SADDLH--PYIRALLTNVLSNCGYQPEDSIRK---VVD 1149
Qy      1153 SVNCLRL---GSKKCTLLILRLQMKRELSVDEILGPTLEIYGLVQADQOLMEKTKA 1208
Db      1150 SINMLRTSIANGDIKEPLDVVTDNTNNSNF-SLEETNASVDMDES-----EKIEN 1199
Qy      1209 KVFSAPIIVLQMKEMKXSDIPQYSQV-----LWVCETL-QEVIATL 1249
Db      1200 GTFQAVL---LKMENDIVYSQFEMLSKQIELTVQNASYEGDLMONLCLFLYTRDDVQM 1255
Qy      1250 FDQTRHSIALGASATEDKSDMETDCSRSRHRDQRGV-----CVLG-LH 1292
Db      1256 NADVQFVTSGLIYEIKKLAGSTQCKAKNQGLGPAVGLLSLEFISHDPTQNKAKNCLVSLBN 1315

```

QY 1293 LAKEI CEVDEGDSWLTQTRRLPILPTLLTLEVS-----RMKONLHPTAHLHLLTLAR 1349  
 DB 1316 FPLELHLSLNNG---LPBDSKHTSIVCIIYLLLEVLADSKKDEFEFNSDCSLKLT----- 1368  
 QY 1350 TQOGATAVAGAGITQSICLPILSYQVLTSTNGTAQTPSARSKSLDAPSPGVRSLMSIME 1409  
 DB 1369 -----TDGALITVDLASQKH-----MSSVI 1388  
 QY 1410 QLLKTLRNLPLBEALDPVGNHOERTLOCLANVRTYQSLACLEADHTVGFILQLSNPK 1469  
 DB 1389 TLLSLSNAN-----LGV-----VSAFRVAVLLTTSASEMHT-----FVKLSG----- 1426  
 QY 1470 WHFHLPOLMRDIOVNLGYLCACTSLH-----SRKK-----LQ 1503  
 DB 1427 ---LPSLFKAMRACSGF-----CNSLHIFRISILRLLPEPEVELMMFDDLVNIFKQ 1478  
 QY 1504 HYLONKNGDGLPSAVAQVQRPSPASASAPSSSKQPAADTEASEQALHTVQYGLKLIS 1563  
 DB 1479 GRARKTEHLGFLRANAEMVLSPEC-----VCOILLDQSLDLAEY 1508  
 QY 1564 KTLAALRHFTPD-----VCQILLDQSLDLAEY 1590  
 DB 1509 KDCCLVGHFTPESEHYIELKESLPVLOQNGOTLDPSKEQMSVYVSFLDELMDLTE- 1567  
 QY 1591 NFLFALSTPTPTPDESEVAPSFQTLTATVNALNMLGEL 1628  
 DB 1568 ---TROPSDRSPNSEFTPENDSLVMYVFLQCTEL 1601  
  
 RESULT 15  
 T00637  
 hypochetrical protein H\_GS541B18.1 - human (fragment)  
 C:Species: Homo sapiens (man)  
 C:Date: 01-Feb-1999 #sequence\_revision 01-Feb-1999 #text\_change 05-Nov-1999  
 C:Accession: T00637  
 R:Hints: K.; Sutterer, C.; Becker, M.; Hawkins, M.  
 submitted to the EMBL Data Library, January 1998  
 A:Description: The sequence of H. sapiens BAC clone GS541B18.  
 A:Reference number: Z14196  
 A:Accession: T00637  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Release: 1-1922 <HIN>  
 A:Cross-references: EMBL:AC004013; NID:G2781380; PIDN:AAB96867.1; PID:G2781381  
 C:Genetics:  
 A:Map position: 7621  
 A:Note: Intron positions not resolved (incomplete sequence)  
 A:Note: H\_GS541B18.1  
  
 Query Match 1.7%; Score 151; DB 2; Length 1922;  
 Best Local Similarity 18.7%; Pred. No. 0.23;  
 Matches 304; Conservative 223; Mismatches 561; Indels 534; Gaps 69;

QY 338 NPEETSSVVRKIGSTAIQNVFOYLTRLLOSLSAGNDCTTACMCVGLSFLVLSLE 397  
 DB 544 --SQBCEVKRRREIEIOLN--EVIETKLOELANIIGK-----TSMN 581  
 QY 398 LHTLGNODIITDPACEVLADPSLPFLPWGTEPTSGILIDSVCGMFPHLSPLLOLIRA 457  
 DB 582 AHSISEADSLKHQDVLVIAEKL-ALBQOYETANBEETPKNVKATFNPKNQLTOELFS 640  
 QY 458 LVSGSKTAKKYVSF-----LDKMSFYNELYHKNH-DVISHEDGTLMRQTPKLLYP 508  
 DB 641 LKRESEVEKIQSIPENSVMVADHLS-----KDKPELEVVLTEDA-----LKS 684  
 QY 509 LGGGTNRI--PGSTQVWMLDRAYLVKREYSVSWTLPTCELEMLHVVSTADVIOHC 566  
 DB 685 LENOTYKSPBENKSGIINLETR-----LLQJEST----- 715  
 QY 567 QRVKPIIDLVAHKVISTDLSIADCLPI-----TSRIYMLQRLTTVISPVPVIA 616  
 DB 716 -----VSAKDLELTQCYKQKDMQEOGFETMLQKKIVLQKIVE--KVAA 761  
 QY 617 SCVNCITVLAARNPAAKWTDLRHGFLPVAHPVSSLSQMSABGMNAGYGNILNMSQ 676  
 DB 762 ALVSQIOLEAVQBYAKFCOD-----NOTISSEPERT--NIQNLNOL 800  
 QY 677 PQGEYGTIAFLRLITLVKQ-----LGSTQSGVPCVMFVLKEMLPYHKRY 727  
 DB 801 REDLSGDISALTRISELSQVVENHTSILEQOYELAKVLEKXKLLLEQKLEB 860  
 QY 728 NSHGVR--QIGCLILHAILINLCHETDLSHTPSLOFLCTGSLA-----YTEA 777  
 DB 861 NEKKOREKEKRSPODEVLTCTTELPHSNE-ESGFNELEALRAESVATKAELASYEK 919  
 QY 778 GQTVINIMGVDTIDMVAQAQRSDAQEGCGQGLIKTVKLAFTNNVIRLKPSSNV 837  
 DB 920 AEKLV-----OBEILLVKE-----TN----- 933  
 QY 838 VSPLEQALSQ--HGAHNNLIVAK-----YIYHGDALPRLAIQLKRLATVAPMS 889  
 DB 934 MTSLOKLSQVRHIALEAKKLSILEKEDTEVQESKACMFPLPKLSKSLASQDGT 993  
 QY 890 VYACLGDAAI--RDATFLQSKIEDMRIRKIVMLEPLTVAVETOPGLIEFLNLEVK 946  
 DB 994 LKISSNQTPQIIVKXAGIQINLQSECSSEVEIISQFTEKIEKQ----- 1041  
 QY 947 DSGDSKEPSLGMWSCIHAVLELDSQ-----QDRVYCCPRLHRAATAFLHLM 996  
 DB 1042 -----LHAA-ETLDMSRHSISETTLKREHYAVAVQL-KEBCGTLKAVI 1083  
 QY 997 QDRDSAMLVARTKPKWENLTSPFLGTLSPS-----ETSEPSILETCALIMKIC 1048  
 DB 1084 QCLRSKGLRAVHNEMOVLSL-----TESPYSDGDEHDSIQVSEPLERKAYINVI-- 1137  
 QY 1049 LEIYVVYKSLDQSLQDTLKKPSIEKRFAYWSGYKSLAVHVAETBSSGCTSLLEYOMLV 1108  
 DB 1138 -----SSLKDLITKQLORE-----AEVYDSS-----QSHESF 1165  
 QY 1109 SAWRMILLIATTAADIMHLTDSVVRQPLFVLDTGKALLVPAVNCALSGMKTLIL 1168  
 DB 1166 SDRKELLIAL-----QOVPLE--ERSVILAAFRLELTALGTTDAVGL 1207  
 QY 1169 ILIRQWRRELSGVDEILGPLTEILEGV-LOADQOLMEKTKAKVPSAFITVL--QMKEMV 1225  
 DB 1208 NCLEQRIOE-----QGVETQAAMECLOKDRRLSLBIQALHQAQNGRKI 1252  
 QY 1226 -----SDIPQSOVLAVNCETLOEVLALFQUTHSLALSGATEDKXSMETDDCSRR 1278  
 DB 1253 TLKRESEKSPGELLEVIN-OOQOSQMLEM--QVEISMKRATTELQELSEK----- 1304  
 QY 1279 HRDQDGVCVLGLHAKELCEVEDGDSWLTQTRRLPILPTLLTLEVSJRMK-ONLHPT 1337  
 DB 1305 -----MVVAELSELAAQ-----TKLELETTLAQKHKLK 1333

Gy	1338	EA-----THTLL-TIARTQOGARAVAGAGITOSTICRPLSVQLSTNGTAQT	1384
Dd	1334	ELEAFRELVKOTBDEVHLLNDPTLASEQKKSRHELOALEKEKAUKL-----GRS	1380
Gy	1385	PSASRKSLDAPSWPBGVYRLSMWSLMEOILKTIRYNFLPEALDPFGVHOERTLOCIN-AVRT	1443
Dd	1381	EERDKXELE-----DLKFSLSEQKKRNULNL-----LEQQQLINESQOK	1422
Gy	1444	VQSIALCEADHTVGFTIQLSNFMKEWHFHLPOLMRDIQNVLGYLCQACTYLHSRKMTQ	1503
Dd	1423	TESQRMLTDA-----QLSEBOGR-NLEIQLVLLSESKVAI---REMSSTLPRERELH	1469
Gy	1504	HYLONKNKGGLPSAVALGRCORPPSAASAAPSASSKOPADTEASEOQALHTYQUGLYKLIS	1563
Dd	1470	AQLQSSDOGTG-----QSRPF-----LPSPDLIKBLQKOLEBKGRSHIVELLN	1510
Gy	1564	KT	1565
Dd	1511	ET	1512

```

RESULT 16
152300
gi|ntlin - human
N:Alternate names: gcp372
C:Species: Homo sapiens (man)
C:Date: 01-Nov-1996 #sequence_revision 01-Nov-1996 #text_change 09-Jul-2004
C:Accession: 152300
R:|Sonda, M.; Misum, Y.; Fujiwara, T.; Nishio, M.; Ikehara, Y.
Biochem. Biophys. Res. Commun. 205, 1399-1408, 1994
A:Title: Molecular cloning and sequence analysis of a human 372-kDa protein localized in
A:Reference number: 152300, MUD:95100974; PMID:7802676
A:Accession: 152300
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-3225 <RES>
C:Cross-references: UNIPROT:Q14789; GB:D25542; NID:g662389; PIDN:BAA05025.1; PID:g808865
C:Superfamily: giantin

Query Match          1.7%; Score 151; DB 2; Length 3225;
Beet Local Similarity 18.0%; Pred. No. 0.53;
Matches 332; Conservative 273; Mismatches 661; Indels 574; Gaps 76;

Cy 34 SQT-EALINKRRRLLEGLSTYKPPSSAEKVANKVAPLKEIGRIKPLGIDBEQ 92
Db 657 SQILEELNFH-----KAOEIKRNLDKANEISNLNQIIE- 693
Cy 93 SVQLQLCYLGEDYGRTRD--SVKTVLDERQSQAIIKINDYYEERTCILRCYHLIT 149
Db 694 -----FKGNADNNSSAFALSEERDQ-LLSQVRE-----LSMVT 726
Cy 150 YFODERHPVEYADCVDLKELKELYSKYROOFEEELYKEATFTWETHGNLTERQVSFW 209
Db 727 ELRAQ-----VKOLEMNLLAEARQR-----RDYESQTHDLILTE-QIHSLSI 769
Cy 210 QCLREQSMLEETIELFYAYFEMAPSDDLVLTMKEQGFGSGRQTNRHILVDTMDPF--VD 267
Db 770 EA-KSKDVKIEVL-----QNEILDVQLQFEGSQETTLIRLSIQQLONKSESEVLGA 818
Cy 268 RIGYSAILIVEGMIDIESLHKCALDDRELHQFADGLICQ---DMDCMLTFPGDIPIHA 324
Db 819 RVRRHIS-----KYSELSQL--SQELETTKKDOLLERKRIVELLQOTI----- 862
Cy 325 PVLAMALLRHTLNPEETSIVRKIGTALQINLVFOYLTIRLIQSASGNDCTSTMAC 384
Db 863 -----EEKDOOV-----TELSSMTKKNVQLNEKFSLG----- 891
Cy 385 VYGLLSFYLTSLEHTLNGOODIIDTACEVLADPSLPFLFMGTBETSGIGIILDSVCMF 444
Db 892 -----VEITTIKEQNLILSRABEAKKE-----QVEDNEVESGGLKNYDE----- 931
Cy 445 PHLLSPILLQLRALVSGKSTAKKTVSFIDKGSFNVELYKHKPHDYVISHEDGTLNR-RQTP 503

```

Db	932	----	MS-	-----	AAQISKEBHQHFDLLKKENBERRKKLQAALINRELLQVR	SRL	979
Qy	504	KULYPLGGQNTNLRIPGTV--	GOYMLDDRAYLVRWEYSYSWTLPTCEIML--	HVVS	TA	560	
Db	980	ELANLKDSEKKEIPLSETERGEVEBDE--	NKEYSKCVTSKCEIIEIYKOTISEK			1035	
Qy	561	DV-IQHQR--	-----	YKRI--	-----	IDLXKVI	STDL
Db	1036	EVELQHRKDLKEEKLAAEOPQALVKOMQNTLODKNTQIDLLQAEISENOAIIOKL--	IT	1093			
Qy	595	SRI-----	YMLQRLFTVISP	PVDV	IASVCNCLTVLAANRPAAKWTDLRHGFLP	VFVAH	648
Db	1094	SNTASDGDSDVAALKEIVISIP-----	CTGSSEHK--	-----	PELEE	1130	
Qy	649	PVSSLSQ-----	MISAE	GM-----	NAGCYNLMNS	EQPOG	679
Db	1131	KILAEKKEQLOQKKLOEALTSRAIKLKAQEKERHREELKQOKDQDYNRLOEPOEDQSK		1190			
Qy	680	E---YGTIATFLR--ITTLXKGGQSGSQS---	GLVPCW	FVLKEMLP	STHKRRYNSHG	732	
Db	1191	ENENIGOLROLOQVRESIDGKLPSTDOOSCSSTGLBEPFLKATEQHR--	TOPV	1245			
Qy	733	REQICILBELHAIILMCHETDHSHTSLOPCLISLAYEAGOTVINIMIGIVDTI		792			
Db	1246	LES-----	NLCEDWPSHSDA	SLQ-----	GGTSVAQI-----	1273	
Qy	793	DMVNAAPRSDGAEQOGOLITTVKLAFSVTNNVIRLKPNSVVS	PRLQALS	OHGANG	852		
Db	1274	-----	KAQKEI	EA-----	KVELBELKVS	STTSSELTKKSBEV	QLOEQINKQ---
Qy	853	NNLIAVLAKYIYKHGDALPRLAIQILKRLATVAPMSVYACLGNDAA	IDAFLTR	QSK	912		
Db	1317	-----	LELES	KTIVS	HEA-----	EVAHES	LOQK
Qy	913	IEDMRKIMLEPFLTVAETOPGILEFLMLEVWDGDSGSKERSLGWSSCLH	AVLEIDS	972			
Db	1340	LESSOLOAGIEHUR--	ELQPKDEL-----	OKLISK	1369		
Qy	973	QOQDRYMCPLRHLRAAIFLHALMWDRRDSAMLTRKPKFWENTLSP	PLFGT	LS	PSSETS	1032	
Db	1370	KEED-----	VSLSQ	QISE--	KEAAL	TRIGI	ELIEQEDLI
Qy	1033	EPST---LETCALIMKICILEIYVYVXGSLDSL-----	KDTLXK--	FSIEK	RPAYMS	1080	
Db	1417	DERIKQLOVEICEKQKPEIEIGESSRAKQOIQRLQALILSRKAL	KENSLO	EBELSLAR	1476		
Qy	1081	GYVXSLAVHVAETGSSCTSLIEQMLVSA	RMILLIAT	HAIDIMH	LD	VSVARQ-----	1135
Db	1477	GTERLTYSKLADVSQVSAQNKEDVYLG--	RALLQ	OEERDKLIT	EMDRSL	LENOS	ISS
Qy	1136	-----	LFLADV	DGK	ALLVPA	SVNC	LTG
Db	1535	CESLKALBEGITEXEKLV-----	KEISL	SSKIA	EBSTEM	QENKELQ	KEYEILLQS
Qy	1174	WKRELSVDEILGFLTEILEGVLQADQOOLMEKTKAVFS	AFIVLQ	NKEMK	VSDI	POYSQ	1233
Db	1588	YENVSNEAERI-----	QHVVEA	VRQEKQ	ELIGKLR	STANKKETEKQ	QOEAE--
Qy	1234	LVLAVCETLOEVAIALDQTRHSLALGSATBD--	KDSMEMD	DCGR	SHRORQ	SGVCLGL	1291
Db	1643	KRRFPANSKQOKLILEEBENDRLAEVHPAGDAKCEMETL	SSNA	MEKELE	VKNMEYE	1702	
Qy	1292	HLAKELCEVDDGDSMWQVTRERPLRPLTLTTELEVS	LRM	QNTLH	FEATATHL	LLTLARQ	1351
Db	1703	TLSKKFGSIMEKRSLSBEVQ-----	DLXHQ	IGBNS	KQANLE	ATEK--	HDQNTVN--TE
Qy	1352	QGATAVAGAGITQSI	CLPLLS	VOLSTNG	TAQTPAS	RKSLDAP	SWGVRL
Db	1754	EGTOSIPETBEQ-----	DSLMS	TRP	CS	ESVPA	KAN-----
Qy	1412	LKTRIRYPLPALD-----	PVG	QOER-----	TQ	CNA	RTQOS-----
Db	1802	-----	NNVL--	QO	IDL	KER	LAG
Qy	1802	-----	NNVL--	QO	IDL	KER	LAG

Oy 1451 EADHTVFILO-----LSNFK-----EWHFHLPOLMRDIOVNLCOACSTL 1495  
 Db 1654 QEBYTKMLLQOIOELSRRTKKEABEEDDLSEELMNOALNINSIGNYCODVTA 1913  
 Oy 1496 LHSRMTLOHYLON--KXNGDGLPSAVAQVORPPSAAS-----AABSSKOPAADEA 1545  
 Db 1914 QIKELLESEMKNLKKCVSELEEEKOQLVKEKTVSEIRKEYLEKIGAOKEPQNKSHA 1973  
 Oy 1546 SEQALHTVOYGLKILSKITLAALRHFTPDVCOILLDSLDLAEVNFPLFALSTPTPTDS 1605  
 Db 1974 KELOELLEKEQOEYVQLOK-----DCIRY-----QE 1999  
 Oy 1606 EVASFSGTLLATVVALMMLGELDKKEPLTOAVGLSTOAG--TETLSLIMFTWNCFY 1664  
 Db 2000 KISALERVYKALFEYQTSOQDLTETKENLAQVHRKKAQAEIASFVLLDDTQ----- 2054  
 Oy 1665 LLISQAMRYLRDPAVHPRDKQRMKOELSELSTLLISLSR 1704  
 Db 2055 ---SEARVLADNLKKELOSNKESVSKQMKQEDLER 2091

## RESULT 17

A56539  
 giantin - human  
 N:Alternate names: macrogolin  
 C:Species: Homo sapiens (man)  
 C:Date: 19-Oct-1995 #sequence\_revision 26-Jan-1996 #text\_change 09-Jul-2004  
 C:Accession: A56539, S37536  
 R:Seel19, H.P.; Schranz, P.; Schroeter, H.; Wiemann, C.; Griffiths, G.; Renz, M.  
 Mol. Cell. Biol. 14, 2564-2576, 1994  
 A:Title: Molecular genetic analyses of a 376-kilodalton Golgi complex membrane protein  
 A:Reference number: A56539; MUID:94187728; PMID:7511208  
 A:Accession: A56539  
 A:Molecule type: mRNA  
 A:Residues: 1-3259 <SEE>  
 A:Cross-references: UNIPROT:Q14789; EMBL:X75304; NID:9405714; PIDN:CA63052.1; PID:94057  
 C:Genetic8:  
 A:Gene: GDB:GOLGB1; GCP; GCP371  
 A:Cross-references: GDB:454958  
 A:Map position: 3q13.31-3q13.31  
 C:Superfamily: giantin  
 C:Keywords: coiled-coil; Golgi apparatus; transmembrane protein  
 F:3238-3254/Domain: transmembrane #status predicted <TMN>

Query March 1.7%; Score 151; DB 1; Length 3259;  
 Best Local Similarity 18.0%; Pred. No. 0.54; Mismatches 661; Indels 574; Gaps 78;  
 Matches 332; Conservative 273;

Oy 34 SGI-EAELNKIMRRLLEGLSYKPPSPSAEKVANKDVASPLKELGLRISKFLGDEEQ 92  
 Db 691 SOLLELELNF-----KAQEIYEKNLDEKAKEISLNQLEE- 727  
 Oy 93 SVQLQCTLOEDYKTRD---SVKTVLODEROQALIKIADYYEERTCILRCVLIHT 149  
 Db 728 -----FKKNMNNSSAFALSEEDQ--LLSQVKE-----LSMVT 760  
 Oy 150 YFDERHRYVEVADCVCKLEKELVSKYRQOFEEELYTEAPTWETHGNLMTEROVSRLFV 209  
 Db 761 ELRQO-----VQOLEMNTLAERQR-----RIDYESQTAHDLNLTB-QIHSLSI 803  
 Oy 210 QCLREGSMLTEIFLYVAVFMAPSDLLVTKMFEKQGFGRQTRNLVBTMDPF--VD 267  
 Db 804 EA-KSKDKIVEL-----QNELDVDLOFSEOSTLIRLSQSQLOKNESVLSGAE 852  
 Oy 268 RIGFSALILVSGMDISLHCALDRRLHQFADGLICQ--DMCMLTTFGDPIHHA 324  
 Db 853 RVRHIS-----KVELSQAL--SQKEIITVMDQILKEKKDVEITLOQTI----- 896  
 Oy 325 PVLLAMALLRHTLNPBETSSVVRKIGTAIQANFOYVTRRLQSLASGNDCTTACMC 384  
 Db 897 -----EEKDOQV-----TEISFSMTERKVVQNEKEPSIG----- 925

Oy 385 VYGLLSFVLSLIELHTLGNODIIDTACEVLADSLPELFWGTPEPTSGLIILDSVGMF 444  
 Db 926 -----VEKIKLEQNLISRAEAKKE-----QVEDNEVSSGLKQNTYB----- 965  
 Oy 445 PHLLSPILLIRALVSKSTAKKYVSFLDKMSFYNELYKHPHDVISHEGTLWR-RQTP 503  
 Db 966 ---MSP-----AQIKSEILOHEFDLKKENEORRKLQALINRKEILORVSRLE 1013  
 Oy 504 KLYELPGGQTNLRIPQSTV--GQVMLDDRAVIVAMWESYSWTLFTQTEIMLT--HVSTA 560  
 Db 1014 EELANLDESKEIPLSETERGEVEEDKE---NKEYSEKCVSKCOEITELYQOTISEK 1069  
 Oy 561 DV-IQHCOR-----VKPI-----IDLVHKVISTDIADCLPIT 594  
 Db 1070 EVELQHIRKOLEEKLAEBOFQALVKQMNQTLQDKTNQIDLLQAEISBNQAIQKL--IT 1127  
 Oy 595 SRI-----YMLQRLTTVISPPVDVIASCVCNCLTILANRPKAVTDLRHTGFLPRVAH 648  
 Db 1128 SNTDASDGDSEVALVKEVTVISPP-----CTGSSHHMK-----PELEE 1164  
 Oy 649 PVSLSQ-----MISAEGM-----NAGGYNLNMSBQPG 679  
 Db 1165 KILALEKEKOLOKQLOEALTSRAILKKAQOEKERRHLREELKQOKDDYNRLQOEPDQSK 1224  
 Oy 680 E---YGVTIAPLR-LITLVKQGLSTQSO---GLVPCVFWLKEMLPYHKMRVNSGV 732  
 Db 1225 ENENIGDQLQLOQVRESIDGKLPSIDQESCSSTGLBEPLFKATEQNH-----TQPV 1279  
 Oy 733 REQIGCILBELHAIINLCHETDLSHSHPSLOPLCLISLAYEAGQTVINIMIGVDTI 792  
 Db 1280 LES-----NLCPDWPSHSDASALQ-----GGTSVAQI----- 1307  
 Oy 793 DMWAAQPRSDGAEQGGQGLLITKYVLAFSVTNNVIRLKPNSVNSPRLBOALSHGANG 852  
 Db 1308 ---KAOLKEIEN-----KVELIKVSTTSSELTKKSEVQLOBOQIKQK--- 1350  
 Oy 853 NNILAVLAKYIYKHGDPRLALIQLLKRLATVAPMSVYALGNDAAIRDAFLTRLQSK 912  
 Db 1351 -----LEIESLKVTSHEA-----EVHAEISLQK 1373  
 Oy 913 IEDMRKIMLEPLTVAVETQPGILIEFLVLEVKDSDGSKERSLGMSCLAHVLLEIDS 972  
 Db 1374 LESSQLOIAGLEHR--ELQPKIDEL-----QKLISK 1403  
 Oy 973 QOQDRWYCPPLHRAALFAFLMWRDSDAMLVARTKPKWMENTPLFGTSPSEBTS 1032  
 Db 1404 KEED-----VSYSQGLSF--KEAALTKIQIETIIEQEDLIRALHTQLOEMQAKH 1450  
 Oy 1033 EPSI---LETCALIMKICILEIYVYVKSLSQSL-----KDTLKK-PSIEKRFAYWS 1080  
 Db 1451 DERIKQLOVELCEWKQKPEBIGESRAKQIQRLQALALISRKALKENKSLQBELSLAR 1510  
 Oy 1081 GYVKSIAVHVAETEGSSCTSLLEYQMLVSAKRMILLIATTHADIMHTDSVVRQ----- 1135  
 Db 1511 GTIERLTKSLADVESQVSAONKEDTVLG--RLALLQEBEDKLITEMDRSLLENQSLSS 1568  
 Oy 1136 -----LEFLVDLGTKALLIPASVNCRLSGMCTILL-----ILLRQ 1173  
 Db 1569 CEGSLKALBELTDEKELV-----KEISLSKSKIAESTEMOEKKEILOKEVEILLQS 1621  
 Oy 1174 WKBELSGVDEILQPLTEILEGVLQADQOQMLEKTKAKYVSFIVYLQNKEMKVSIDIPOYSQ 1233  
 Db 1622 YENAVNSAERI-----QHVBAVAVQEKQELYGKLRSTANKKETEKQLOEAE-QEMEBMKE 1676  
 Oy 1234 LVIANVCTLOEBVIALFDQTRHSLALGSALBD--KDSMETDDGSRSHRQROGVCTGL 1291  
 Db 1677 KMRFAKSKQOKILIEEENDRLRAEVHPAGDTAKCEMTLLSSNAKKELELVKAEYE 1736  
 Oy 1292 HLAKEICEVDEDGSMLOVTRRLPILEPTLLTLEVSIRAMQNTLHFTATLHLTLTLARQ 1351  
 Db 1737 TLSKKPQSLMSKEDSLSEBQV-----DLGHQIEDVNSKQANLEATK--HDQNTNV--TE 1787  
 Oy 1352 QGATAVAGAGITQISICLPLLSVYQLSTNGTAQTPSASRKSLLDAPSWPGVYRLSNSIMEQL 1411

Db 1788 EGTSGISGETEQ-----DSLMSSTRPTCSSEVPSAKSAN-----PAVSK-DRSSHDEI 1835  
Qy 1412 LKTRVFLPEALD-----FVGHOER-----TLQCLNAVTVOS-----LAQL 1450  
Db 1836 -----NNYL--QOIQDKERIGLEEKOKNKEFSQTL--NEKNTLLSQSTKXDELKML 1887  
Qy 1451 BEADHTVGFILQ-----LSNFMK-----EMHFLPLQMLRDIVQNGVLCQACSTL 1495  
Db 1888 QEEVTKNNLNLQIQEELSRVTKLKEAEEKDXLEERLMOQLAELNGSIGNYCQDVYDA 1947  
Qy 1496 LHSKMLQHYLQ--KNXGDLPSVAQRVQRPSPAS-----AAPSSKQPADTEA 1545  
Db 1948 QIKNELLESEMKNLKCVCSELEBEKQDYKERTKVESIRKYLEKXQAGAKERGNKSHA 2007  
Qy 1546 SEQALHTVQYGLIKLISKTLAALRHPTDVQCILLDQSLDLAEYNFLPALSFTTPTFDS 1605  
Db 2008 KEIQELKEKQEQEVQLOK-----DCIRY-----QE 2033  
Qy 1606 EVAPSPFTLLATVVALNMLGELDKKEPLTOAVGLSTQAG--TRTLKSLMFTMENCY 1664  
Db 2034 KISLERTVAKLEFVQTESQKDLITKENLAQAVEHRRKQAQELASFKVLLDDTQ----- 2088  
Qy 1665 LLISQAMRYLDRPAVHPDKQRMQOELSELSTLSTLSR 1704  
Db 2089 -----SEARVLAADNLKKELOSNKESVSKQKQKDEDLER 2125

## RESULT 18

Db 87755 protein T21E12.4 (imported) - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 24-May-2001  
C:Accession: D87755  
R:anonymous, The C. elegans Sequencing Consortium.  
S:Title: Genome sequence of the nematode C. elegans: a platform for investigating biology  
Science 283, 2012-2018, 1998  
A:Reference number: A75000; MIMD:9906613; PMID:9851916  
A:Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/projects/C\_elegans/; published extra appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and  
A:Accession: D87755  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-4464 <STO>  
A:Cross-references: GB:chr\_I; PIDN:AA837647.1; PID:g1703597; GSPDB:GN00019; CESP:T21E12.  
C:Genetic8:  
A:Gene: T21E12.4  
A:Map position: 1  
C:Superfamily: dynein heavy chain, cytosolic

Query Match 1.74; Score 151; DB 2; Length 4464;  
Best Local Similarity .17.0%; Pred. No. 0.91;  
Matches 283; Conservative 233; Mismatches 533; Indels 614; Gaps 71;

Qy 346 VRKI-----GGTAIQ-----LNVFQYLTRLIQLSLASGNDCT--STACMCYVGLSLF 391  
Db 148 IRKQTQLEERDSSGTSIQEMTFMLNERALKITISK--RQGEVYTLLEALKC----- 198  
Qy 392 VLTSLELN--TLGNQOD-----IIDTACEVLADPSLPELFWGTEPTSGIILDSVCG 442  
Db 199 ---GKRPHATVGFPSDNLQKLAIVQDYNTLMKEFPLSELVSATDVPK---LMHAYVG 251  
Qy 443 MFPPLLSPLQLPLALVSGSKTAKVVSFLDKMSFYNELVGHKPHDIVSHEDGTLMRQT 502  
Db 252 IFLH-----LKRKSTKYPQLRALVLEAIS-----DLN 281  
Qy 503 PKLILYFGGQTNLIPQSTVGQVMDRAVILVMEVYSSTWTLFCEIEMLLHVNSTADV 562  
Db 282 SQLKLVSSYNNLMKTPLAERFELMSQCALPSKWDDEYDKF-----TALIRD 328  
Qy 563 IOHCORVKEPI-----IDLVHVISTDLADIADCLPITSRIYMLQ-----RLTVVISPP 611  
Db 329 INKKRRDPSLSLMSKVAVVHRLFT-----RLMQIQLQFRKHQGFVIER- 374

Qy 612 VDVIASCNCLTVLAARNPAKVMWTDLHHTGFLPFVAVPSSLSQ-----MT-SAEQNAV 665  
Db 375 -----VLRPVNGSRERQMLDSSG----- 396  
Qy 666 GYGNILMNSQPOGEYVTAFLR-----LITLVKG 697  
Db 397 -----EKSPDQVDAIEVFLKNDVFLDVPDSFAMENAFKRYEDQIGVETALTTLTKS 448  
Qy 698 QLSGOSQGLVPCVMFLKEMLSYHKRRYNSHGVROIGCLIELHAILNLCH--TD 755  
Db 449 QLESRRNS-----NEWFSIFS--RYALFLRIRGAIAYEYOTRLINRYKEDINE 496  
Qy 756 LHSHTSLQFLCISLAYTEAGOTVININGIG-----VTTIMVM 796  
Db 497 LQARFTK-----ARGQGVKIQOTGLPFPSPAKIMIRNYEROLQRYMKRVEDVL 546  
Qy 797 AAQPRS--DGAEGOGQOLIKTVKLAFSVTNNVIRLKPSPNVVSPLQALSOHGAGNN 854  
Db 547 GKQENNVDRQLKADD-----NFKYKLTQMPDEWESVQSQWTLIPNK 593  
Qy 855 LIAV-----LAKYIYHKDPLPRLAQLRLATVAMSV----- 890  
Db 594 ILTVDRVQVDRMQLQKINHSDSVLYK--EVSHLKSMGFRVPLKTVMAHQANQMRPS 652  
Qy 891 -----YACGNDAAIR--DAFLTRLQSKIEDMRKIVMTLEF-----LT 927  
Db 653 ATSLIEARTFPASVNAALASQVDSLASYKDOIQLLEGATLGWDSYKVDYOLKLA 712  
Qy 928 VAVETQGLIEFLNLT-----EYDQSDSGSKESPLGMS 961  
Db 713 ETVTVYQERCELLNVRIYNADLNVLKSKRYDEKTEINLLTSIQKVD--QLSLGNY 769  
Qy 962 CLHVLLELDSQOD-----RYWCPELHRAAIAFLHLMQDRDSAMLVLRKP 1011  
Db 770 NLAQVNTLRLQRTIILARRVEDAIRW-----TLVESQSEVELEERQVVLPTV- 820  
Qy 1012 KFMNLSPLFGT-----LSPPSTSRPSTLE-----TCALMKIKICLEYVYVNGSL 1059  
Db 821 ---GNVAVDLCMTAQTLYISPTRETEKLEQLYEMHSVTAQMRISGKRFQVMMDEI 877  
Qy 1102 RMLLIATTHADIMHLDTSVVRQLFLVDLDGYTALLIVPASV----- 1155  
Db 930 QLFEMLGTSLSKMMKTLMERKGRLVPTQDTRK--VIFPVSVEYGAQOKILPKYDYWH 987  
Qy 1156 ---CLRQSM---KCTILLILIRQWR--ELGSYDEILGPTLELBEVLOADQOLMERTK 1207  
Db 988 KEMLVKFGAAGVDMQKRFENSVSKRWRLVLEQSYD---GGSSTPTIGLISVQSLKQTK 1044  
Qy 1208 AKVPSAFTVLOKEMKVSIDIPQYQVLNVNCEYLQEEVIALFPQTRHSLALGATEDKD 1267  
Db 1045 S-----GQAVDLYRSSQRLN-----QQRXQFP----- 1068  
Qy 1268 SMETDSCRSRHRDQDQVCVGLHIAKELCEVEDDGSMLOVTRRLPILFTLLTLEVS 1327  
Db 1069 -----AOMLYSENVBE--MSAFTEILSLDASIQIQGMN 1101  
Qy 1328 LRMQNHLFTEATIHLLTLARQOGAT-----AVAGAGITQSGICPLSVYQSLSTNGT 1381  
Db 1102 LQTK-----FAQED--ELVEKRYVETLTENKSKPVGAQAPQ--ALNVITAFPAKLNKL 1153  
Qy 1382 AQPSASRK-----SLDAPSPG---VYRLSMSLEOILTKLR--YNFLPEALD-- 1425  
Db 1154 TREBNRKARVALDLDSAHAPSEGDKLTVATBELLAAMDQWALQPVYTGIDAEAKET 1213  
Qy 1426 FVGVOERTLOCLNAV-----RTVQSLACLEADHTVGFILQ-----SNFMKE 1469  
Db 1214 WLSVQPKRIRQSLDELANNQLKQLPVKCRITYKSVYEHQMLHTYQKNNMLVALEKSEALK 1273



Db 1157 LMTVAQAGAAQGLAQTGVNNAVGTLPRLRRSRRLRFLASTFLAE----- 1205  
Qy 1560 KIISKTLALRHPTPDVQCQLLDQSLDAEYNFLFALSTFTPTPDSVADSEFTLLATVN 1619  
Db 1206 PMFAETVRAVAVQLOP-----HYIGNYQASTAPLPEEKIVFSLSAATOE 1251  
Qy 1620 VAL-----NMGLDKKEKPELTQAVGLSTQAEET 1648  
Db 1252 LTLAKIRTMYGKMD-AEILSOHLGISPDHST 1282

RESULT 20  
T14156  
kinesin-related protein - African clawed frog  
C/Species: Xenopus laevis (African clawed frog)  
C/Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 09-Jul-2004  
C/Accession: T14156  
R/Mod: K.W.; Sakowicz, R.; Goldstein, L.S.; Cleveland, D.W.  
A/Title: CENP-E is a plus end-directed kinetochore motor required for metaphase chromosome  
A/Reference number: Z17893; MUID:98028574; PMID:9363944  
A/Accession: T14156  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Residues: 1-2954 <MOO>  
A/Cross-references: UNIPROT:O42263; EMBL:AF027728; NID:g2586070; PID:g2586071; PIDN:AA6C  
C/Genetic: XCENP-E  
C/Superfamily: centromere protein E; kinesin motor domain homology

Query Match 1.6%; Score 147; DB 2; Length 2954;

Best Local Similarity 17.1%; Pred. No. 0.88; Mismatches 474; Indels 508; Gaps 57;

Matches 244; Conservative 202; Mismatches 474; Indels 508; Gaps 57;

Qy 557 VSTADVYQHCORVPII-----DLVHKVISTDLIADCLPI-----TSRI 597  
Db 1 MSBGDAVKCVVRPFLQREGDQANLQKAGNTTIGVDGTSFNDVRVNSHESISQI 60  
Qy 598 YMLQRLTVISPPVDVIASCV---NCLTVLAARNPAKWTDLRHGFLPFAHVPVSL 653  
Db 61 YQEI-----AVPIIRALQGYNGTIFAYGQTSKGTYTMMGTNSLGIIPQALQEV 111  
Qy 654 SCWISAEQNAAGCGNLMSSEPOGEGVTIAFLRLITLVKQL----- 699  
Db 112 FKII-----QELPFRBFLRVSYMEIYNETVXDLLDDRRKKPLEIRE 155  
Qy 700 -----GSTOSQGLVP-CVMFVLK-EMLPYHKRVNSHGVRQIGCLIELLHAIL 748  
Db 156 FNRNVYVADLTBELVMVPEHYIQIKGGEKRRHYGETRKMDHSRS-----HTIF 205  
Qy 749 NLCHETDLHSHRPSLOFLCISLAYTEAGQTVINIGIVTIDWMAAQRSDAGEQ 808  
Db 206 RMIVESHDRDPTNSN-----CDGA-----VWVSHNLVLDLAGESEASQTAEG- 250  
Qy 809 GQGOGLIKTYKLAHSVNNVIRLKPNS-----VSPLEBALSHOHAGNNLIAVLAKY 862  
Db 251 -----VRLKEGNNRSFLIQGVIRKLSDGAG-----F 281  
Qy 863 IYHKDPLRLAIQLKR-----LATVAPMSVYACI-----GNDAAAIR----- 902  
Db 282 INNR-DSKLRILONSJGNAKTYITITPVSDFELSTIQFSTAKHVRNTHVNEVL 340  
Qy 903 -DAFLTRLOSKIEDMRKIMLEFLTVAVETQGLIELFLNEVKGSDGSKESLGMW 960  
Db 341 DDEHLTKRYREKIDLLKKOLENL-----SSSETKQAQMAKESHTQ--- 381  
Qy 961 SCHAVLELDSQODRKYWCPPLHRAAIAFLHLMODRPSAMLVLRTPKFE----- 1015  
Db 382 -LLAEIKQHKEREDRW-HLTNIVVASSQSDQR-----VRRKRVTVAPGKIQ 431  
Qy 1016 -----NLTSPLFGTSLSPSETSE-PSILETCALIMKICL----- 1050

Db 432 NSIHASGVSDPDMLSRLPNSFKKAKFSDMPSEFE-----IDDSVCTEFSDFDALSNMDS 487  
Qy 1051 -----IYYVKGSLDQSLKD-----TLK 1068  
Db 488 NGIDAEWNILASKYTHREKTSLHOSMIDFGQISDSVQPHSSKENQOLYLPKDSGMACE 547  
Qy 1069 KFSIERKFAVY-----SGYKSLAVHVAETEGSSCTSLILEQMLVSAVRMLLI 1117  
Db 548 KASFEKITSLOOQLOSKKEEKEKELVOSFELKIALEBQSLSVAKULEMTNS-REHSIN 606  
Qy 1118 ATTHADIMHLDTSVVRRLFDVLDGFKALL-LVPASVNCIRGNS-KCTILLILLRQ 1173  
Db 607 AVQOTDY---EKEVVRKEMSVLDGSGYNASNDLQDSVQDKRLSSSHDECIEHRKMLEQ 663  
Qy 1174 WKRELSGVDEILQPLREILGVLQAO-----QLMETKAKVFAFTVLQMKMKV 1225  
Db 664 ---KIVLEEFIRLNKSSKENDQSSBQPMESIQCEAIMKALBELALMR--- 717  
Qy 1226 SDIPQYQVLYANVCELTQSEVIALFDQTRSLALGSATEDKDSMETDSCRSRRRDQDG 1285  
Db 718 ---NFDNITILE-NETIKREIADL-----ERSL-----KENQETHEFELKEQTEKH 760  
Qy 1286 VCVLGLHLAKELCEVEDGDSWLD-VTRRLPIPLTLLTLEVSU---RMKQNLHFTFAT 1340  
Db 761 EAQL-IHEIGSLKVLVNAEMYNQNLDEDELETYKTLKEQIQLAELRKADNLQKKVRN 819  
Qy 1341 LHLITLARTQOQATVAVAGGITOSTCLPLSYQLSTNGTAQTPSASR----- 1389  
Db 820 FDLVSN-----GDSKELCEELFOLKQSLSDAEAVTRDAQKCSFLRSNLE 866  
Qy 1390 ---KSLDAPSPGYYRLSNLSMEOLKTLRYNF----- 1419  
Db 867 LKEKMBETSNWYOKEXAASLPEKQLETESNYKMAADLOKELQSAFNEINYLINGLAG 926  
Qy 1420 -----LPEALDFVGHQERTLCINAVRV-QSLACL----- 1450  
Db 927 KVPRLDLRVELEKVSFQKQLEKALEKNA-LENEVTCLSEKFLPNEVECKKQISK 985  
Qy 1451 -----EADHTVGFILQSNFMKEHFLPQLMRDLOVNLGYCOACTSLHSRKM 1501  
Db 986 ASEBIMLKQEGEHSASISKQELIMOESBOQLQLDDEVTHROSKQOQEBQYLEMKM 1045  
Qy 1502 -----LOHYLONK-----NGDGLPSAVNQRVQRPASASAPSSKOPADTEAS 1546  
Db 1046 HDLFEKTYINRKBAEDLLBEMENLKGTMESVEKI-----ADTKHELEERTIRD 1094  
Qy 1547 EQQALHTVOYGLKIISKTLAALRHPTPDVQCQLLDQSL-----DLAEYNF 1592  
Db 1095 KEQILHEKTY-----FQAMQTIPIPR-----LSDSLPSKULVEGNSQDPIINDYHN 1143  
Qy 1593 LPALSTFTPTPDSVAPSPGTLATVNAVNLMLGELDKKPELTQAVGLSTQAEGRITLK 1652  
Db 1144 LIAL-----ATERNNINWCELEBNSLKE-----QVIDLNTQ----- 1175  
Qy 1653 SLMTFTENCFFYLLISQAMRYLRDPAVPRDKORMKQEL-SSSLSTLL 1699  
Db 1176 -----LQSLQAGSIEKSDLOKPKQDLBEGVXKLL 1205

RESULT 21  
S48938  
hypothetical protein YHL030w - yeast (Saccharomyces cerevisiae)  
C/Species: Saccharomyces cerevisiae  
C/Date: 02-Dec-1994 #sequence\_revision 02-Dec-1994 #text\_change 09-Jul-2004  
C/Accession: S48938  
R/Favella, T.  
A/Description: The sequence of S. cerevisiae cosmid 9196.  
A/Reference number: S46794  
A/Molecule type: DNA  
A/Residues: 1-1868 <FAV>  
A/Cross-references: UNIPROT:P38737; EMBL:U11583; NID:g2289854; PIDN:AB65042.1; PID:g2289854

C:Genetics:  
A:Gene: SCD:ECM29  
A:Cross-reference: SCD:S0001022; MIPS:YHL030w  
A:Map position: 8L  
C:Keywords: transmembrane protein

Query Match 1.6%; Score 146.5; DB 2; Length 1868;  
Best Local Similarity 18.2%; Pred. No. 0.45; Mismatches 632; Indels 609; Gaps 92;  
Matches 337; Conservative 269; Mismatches 632; Indels 609; Gaps 92;

QY 61 SAEYKANKDVASPLKELGRLSKFLGDEBSQVQLQCYLQEDYKRGTSVKTLQDER 120  
DB STDSTNLSDDAAQMKKPP-----MPEYNEEFITFLQTLVIGTANGRPVKAILDEKI 322  
QY 121 QSQALLIKIADYYEERTCI-LRCVLHLITTPQDERHYRYEYADCYDKLEKELVSKTKQ 179  
DB 323 LS--ILNRSHFATTKACISLISIGL-----HSSEY-----KLSLTLSPFRH 364  
QY 180 QFEELYTEAFETWETHGNLMTERGQSRWFVQCLREQSWLEIFLY--YAFEMAPSDL 236  
DB 365 VAKNTYKLNPA-----SSSPSTPSTCTIVSLINN-----LHAGMKQLQGP-- 409  
QY 237 LVLTQMKRQGGSRQTNRLHVDETMDPEYDR-IGYFSAIILVGMDEISIAKCALDRR 295  
DB 410 -----QTPAFNTAIIQRLQYETLGLILKRPDELVSLSYIRFL-FESL-----KN 454  
QY 296 ELHQFADGLICQMDCLMFLPGDIPHHAPVLLMALLRHTLNBEETSSVVRKIGTALQ 355  
DB 455 DLPGFRSS--IQESLLSLVGHLSILPOOSKIKLK-NILRKLSIDEQO----- 499  
QY 356 LNVQVYLRLLQSLASGNDCTTSTACICVYGLSFYLSLEHTLGNQDIDTACEVL 415  
DB 500 -----REDNNNAVNSIMAL-KVSIKFTNNAEPFH----- 528  
QY 416 ADPELPFLF-WGTEPTSGGLIIDSVCGRP-----HL-- 447  
DB 529 -DPE-ARLFNMGVTRNRPDIIESEFKIQPFWRVNNASINTSATVKTSDLLGSHISE 586  
QY 448 --LSPLQLRLALVS-----GKSTAKKVVSFLDKSPFVNELYGKHGHDVISHDGT 496  
DB 587 TEPPPEPFLQVLIDQDSEAASTTRKSLNNAVFSQCLISNAIYKKTW-VIQDED-- 643  
QY 497 LMRKQTEKLLYPLGGQTNLRIPQGTGQVMLD--RAYLVREWESYSWT----- 544  
DB 644 -MSVRIDKAL-ELDDTVASRVNEMVQG--NMDDIFIRYLLTSLNEPATNSKSGQIAIFP 699  
QY 545 -----LFCSEIMLHVSTADYIOHCQVRPIIDLVAKYI-----STLSTADCLLP 593  
DB 700 YDPIFGSVLLTLNLPVSN-NVLRRLLELVP--DLVHLVIMKQSLSDNDLAVACATIGI 756  
QY 594 TS-----RIYMLLQRLT--TVISPPV-----DVIASCVNCLTVL 625  
DB 757 ISTALADSTHVKRTTKIAQSGTMAETVYASVVRRLYKQTNHIESDSILNLTITL-- 814  
QY 626 AARNPAKVTDLRHTGFLPVAAHVSSLQMSISAGNNAAGYGNLNNSEPOQGEYGVTT 685  
DB 815 -----THLSHPG-----TNKDMILKLVQCVTKFGILLQVSAQERKD----- 850  
QY 686 AFLRLITLVKQGGSGTQSGQVPCWVFLKEMLPSTHKRRYNSHGRQIGCLILELH 745  
DB 851 -FLKRVMDTIQDKLIND-----VTALQTSVYS-- 877  
QY 746 AINTCHETDLHSSHTPSLQFLCISLAYTEAGQTVININGIVDTIDWMAAQRSDGA 805  
DB 878 -----LYSTDLENS-----SLFQKLELTNVSCKN-----DPLFSGESLSV 915  
QY 806 EGGGQGGQLIKITYKLA-FSVTNNVIRLK-PSPNVVSPLEQALSQHGAGNNLAVLAKYI 863  
DB 916 AGKWSKYLKQIDIPFNV--EIMQOKFPATVNTITLD-----1 954  
QY 864 YHNGDALPRL-----AIQLKRLATVAPMSVYACLGDA-----AATPDAFLTRIQ-- 910  
DB 955 FSGCDSTKPSLRKASCIWLSYIQYGLHPVSSKCDIHLRFRPLADDEFIQDSAPR 1014

QY 911 -----SKEDMKIKMILEFL-----TYAVET--QPGLEL-- 939  
DB 1015 GLSLVEIGSDIKESVWKGILKSFTESTAGSASTATGISVSETELFEPGLVLTGCD 1074  
QY 940 -----PLNL--EYKQSDGSKESP--QWMSG-----LHAYL-----ELIDS 972  
DB 1075 GSISTYDILNLASEVDDPALVYKFNLSLAASSLMSRKRIAGLGAIMSKSLIEELLK 1134  
QY 973 QQDPRWCPLLRRAIAFLHALMODRRDSAMLVLRTPKFW-----ENLTPFLF-- 1022  
DB 1135 DQ-----TAKLIPKLYRFPDQVAVSRMTDIIWMTLIPESLITSLYENDI 1183  
QY 1023 -----GTSPSESPSILETCALIMK-----ICLEIYVVGSLDQSLKDT 1066  
DB 1184 LDELICGMANKRVRERAS--TSALLQLIQSOPQOEKPSKMLKIWMARFTWD-DIKDS 1239  
QY 1067 LKFSIEKRPAYSGVYKSLAVHVAETEGSSCTSLLEY--QMLVSAM-----RML 1114  
DB 1240 VREVGK-----FTVLAKILARSIDVEKGVNPTKSEILDNIIFPLMGPHGLNSDAEVR 1295  
QY 1115 LIAATVHADIMLTDVVR--RQLFLVDLGTKALLVPASVNCRLSGMKCTLLIL 1171  
DB 1296 NPAITLIDLVKSPGAIKFPKLYDFI--TLFSIEPQVINYALNANYNIDANVI 1353  
QY 1172 ROKKRELGSVDLGLPTEILEGVL-QADQQLMEK-----TKAYFSAFT 1216  
DB 1354 -DTQRKNGVTN--SPLFQTEKLIINNSDCMMEIINVIKASRKSGVGLPSKVASLVI 1409  
QY 1217 VLQMKEMKVSIDIPOYSQVLNVNCEITLOE-----VIALPQTRHSLA 1258  
DB 1410 IILVKRYSI-EMKPYGKILKVLTFMEDRNESVNIAFASMGYLRKVSALDKCIKSEK 1468  
QY 1259 LGSATEDKQSMETD-----DCSRSRHRDQGVCVGLHLAKELCEVEDGD----- 1305  
DB 1469 LITKYPEPTSTENKKRVGTALDISILNVAKSEBDNVAVEMPLIFIACN-DEKQDLETLY 1527  
QY 1306 --SMLQV-----TRRLPILPTELTLEVSILRKQKULHFEATLHLLTLARPOGATN 1357  
DB 1528 NKTWTAASSGACTVCL-YLPEILANVCNLIKND-----FSIRKCAKSVIQ 1574  
QY 1358 AGAGITQSLCP-LLSVYOLSTNGTAQTPSARSKSLDABSPWG--VYRLSMSLMOQLK 1413  
DB 1575 LCGINDSIPYQIVKLPDI-----SREALSGRSMQCKEHIIVALVSLTEKFSQ 1623  
QY 1414 TLR-VNPLPBAIDFVGVRQERTQCLNAYRTVQSLACLEADHTVGFIIQLSPKEMHR 1472  
DB 1624 TVADNNDLQESINHV-MYTEVSRKSKKYYKXI----- 1654  
QY 1473 HLPLMRDIOVNLGVLCQACTSLHSRKMLOHLYLQKNGDGLPSAVALQVRQPPASAA 1532  
DB 1655 -LPLVARYINVPQE--EITPLIEKAKEMIRLIGSESDS-BGSLKQ-----TSD 1701  
QY 1533 PSSSKQPADTEASEQALHTVQ-----YGLKILSKTLAALNHTPPDVC-----QI 1579  
DB 1702 ESTIKRIKENTETIQSKSENIENEYVNLKV-----SVDICNNKSRYPWN 1750  
QY 1580 LIDQSLDLAENVFLPALTTPPDESEVAPSGTILLATVVALNMG 1626  
DB 1751 LLEFIID-EIAYLF-----HNDRIIHWTQTOLAASEIGISIVG 1787

RESULT 22  
T32550  
hypothetical protein f33d4.2a - Caenorhabditis elegans  
C/Specties: Caenorhabditis elegans  
C/Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 21-Jan-2000  
C/Accession: T32550  
R/Johnson, D.; Stellyes, L.  
submitted to the EMBL Data Library, December 1997  
A/Description: The sequence of C. elegans cosmid f33d4.  
A/Reference number: Z21190  
A/Accession: T32550



A>Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-2848 <JOH>  
 A:Cross-references: EMBL:AF036702; PDB:1AB8374.1; GSPDB:GN00022; CESP:F33D4.2a  
 A:Experimental source: strain Bristol N2; clone F33D4  
 C/Genetics:  
 A:Gene: CESP:F33D4.2a  
 A:Map position: 4  
 A:Initons: 48/2; 117/3; 167/3; 263/3; 316/3; 362/1; 378/2; 396/1; 436/3; 509/3; 658/3; 7  
 /1; 2324/3; 2430/1; 2492/3; 2597/1; 2823/3  
 C:Superfamily: Inositol-trisphosphate receptor

Query Match 1.6%; Score 145.5; DB 2; Length 2848;  
 Best local similarity 17.0%; Pred. No. 1;  
 Matches 325; Conservative 264; Mismatches 653; Indels 669; Gaps 84;

9 SVLSFCSSRELWTI--LGRSALRELSQIEALNKWRRLLEGLSYKKPPSSAEKVK 66  
 DB 851 SPMSAIHARLWMSIPENVNVSTYESVS-VEAYSDGSRMRIGEGI-----AHKVL 899  
 QY 67 ANKVAAPLKEGLRISKFLGLDEOSVQLQCYLQEDYRGRSVKTVLQDEROSQALI 126  
 DB 900 ATVETVY---WGLRNO---SMEERQSV-----NSSKLTVEIVNLAKL- 936  
 QY 127 LKIDAYVEERTCILRCVHLHLYFQD---ERHPRYEYADCVDKLEKELV-SKRYQOF 181  
 DB 937 ---AQFHFSPNDLQLTQNLALINEGPATEQVPSHRAVNAIRNMSKMRGKNENS 993  
 QY 182 BELYKTEAPWETHG-----NLMTROYSR--WPVQCLRESMLLEIIFLYYAYPEM 231  
 DB 994 KDLAKTSPVAEEAGRTKEGRALNVKTLIAEILQFVMDVRDRIYRTMALSPKYNF-- 1051  
 QY 232 APSDLVLTKMFKKQGG--FGSRQTNRHLYDVTMPFVDRIQYFALLIVEGMDLESLHKC 289  
 DB 1052 -PC-----DGLSMHSASINERASELYDAIYRSSGH----- 1083  
 QY 290 ALDRREHQAPOGDLICQMDCLMTFGDIPHAAPVLLAWALLRHLNPESTSVYRKI 349  
 DB 1084 ---ELHIDGRGQGLLAI--LLOMTMSDVPRLTSTALK-VPRRHFTQYELLEDKQV 1135  
 QY 350 GGTAIQLVNFOYLTRLIQSLASGNDCTSTACMCVYGLLS--FVLTSL--ELHTLG 402  
 DB 1136 -----QLTVSNNDVEN-----YQIDRDIFILKILTEKSELWVHG 1170  
 QY 403 NQODIITACEV-----LADPSL--PELFWGTEPTSGIILIDSVCGMPPHLSL 450  
 DB 1171 DRHHSIDTK-EVDEKERTTEHLLDHDLSKPRAPDSGSMELAAVLANE--HPSIRNE 1226  
 QY 451 LLOLLRALV---SGKSTAKKVSFLDKMSFYNELYKKHPHVDISHEDGTLWRQTPKLY 507  
 DB 1227 CLQALNLLIKQDNDAAVALQELSDK-----APLIAY 1259  
 QY 508 PLGGQTNLRI-----PGSTVQVMLDRAVLYRWMEYSYSWTLFTCEIEMLLHV 556  
 DB 1260 PLIKQMLVRLGCMYCRKGDPRKPTMNOQLTKMRYV-----EVLLEF 1301  
 QY 557 VSTADVIOHCORVPIIDLVKIVSTDLSTADCLPITSRTYMLQRLTVISPPVDVIA 616  
 DB 1302 ISVPHDKHDHMKMLITLSHEFLR--SFCKTKENOSRLY----- 1340  
 QY 617 SCVNCITVLAARNPAKWTDLRHTGFLPFVAHPVSSISOMISAGMNAGVGNLNMNSQ 676  
 DB 1341 ---KFIYEDAKEGMLRVETIE----- 1360  
 QY 677 PQGEGVTIAFLRLITLVKQGLSGTOSQGLVPCWFMVLKEMLSYHKWRYNSHGVEQI 736  
 DB 1361 ---EVGTLVAIFR-----NNRELASNVPEELIAHIVGLIE-----HNSRNP 1399  
 QY 737 GCULLEIHLALNCHETDLSHTPSLQFLCISLAVTEAGQTVINIMIGVDTIDMW 796  
 DB 1400 ---FLELLQALV-CVYDEKIESGQEKVANEICAAS--DEVROLVVD--NASFELNAM 1450  
 QY 797 AAGRSDGAGGQGGOL-----LIKTVLAFSVNNVNIPLKPSNVVSPLEQALSQGH 851

DB 1451 KDEKSKGRSSDSRRKLYHIELVRLAMCTRGNGTELKASQI--PMD----- 1499  
 QY 852 GNNLIATLAKTIYHKHPALPRLAIQLLKLATAPMSVYACGLNDAAALRDALTR--- 908  
 DB 1500 ---HIVRYVT-----AKQCLVEKYTVYQLLHCHYIDTDAEMKDAKTEYVD 1543  
 QY 909 ---LOSKEDEMRKIMILEFTVAVERQGLIELFLNLEVNDGSDGSKFSIGMSSCLHAV 966  
 DB 1544 HILNMLLED--IRSLREKLT-----GATATLEHTIC-HTV 1577  
 QY 967 LELIDSOQOQRYWCPRLHRAAIAFLHALWQDRDSAMLYLRTKPKFWMITSPDFGTL 1026  
 DB 1578 TEVL-----IKFEPAPYSALQAAVVDVHHKKFTSE----- 1608  
 QY 1027 PPSSETSPLTECALIMKIICLETIYVVGSLDQ-----LKOTLKFF----- 1070  
 DB 1609 ---VLELTYLEKGLRGSKSSSNWYRVACIRKLTGMAEHN 1648  
 QY 1071 ---SIEKRF-----AYMSGVYKSL-----AVHVAETEG 1095  
 DB 1649 ITLPATLAGPOMSGQISVRQWQOASSAKWIGIGKRLNONTLNPGRHLYGTSNTEH 1708  
 QY 1096 SSCTSLIEYOMLYSAMRMLLIATTHA-----DINH-----LTDVVVRQLF 1137  
 DB 1709 TSANVVTGYHMMIGEKFYL--HPLHABGSVLVEVHLTELEFPESALRDQARGGV 1766  
 QY 1138 LDVIDGTKALLVPASVNCIRPLSGMKCTLLILIRQKRELGSVDILGLPT---ELE 1193  
 DB 1767 AKLIQCKTLMONKODNICARVLOTLCKMCDCTKQ-----LTHQGHRYVN 1812  
 QY 1194 GYLQADQOLMEKTKAKYF---SAFITYLOMKEMVSDIPQVSQVLNVCELTQSEVIALF 1250  
 DB 1813 SSVEKGOQLQOLLQRFGGHNNHPPLDROQSKTGE-----VIEAVEKKEETW 1862  
 QY 1251 DQTHSLALGSATEDKDSME--TDCCSRSHRDQDVCVYLGLLAKELCEVEDGDSWL 1308  
 DB 1863 SQERDVLAIQKUNDASDLVDIIMEPSRE--IFKAIHLARALL--HEGNDKV 1915  
 QY 1309 QVTRRLPIPLTLTTEVLSLRMKQ-NLHFTBATLHLLTLARTQO----- 1352  
 DB 1916 QHS-----FYRMKQKQDIH--EPFEKALITITQNRNLYKSDMWSKSDSKP 1959  
 QY 1353 ---GATAVAG--GITQICPLIS--YQIYSTNGTAQTPSAS-----RKS 1391  
 DB 1960 KVSSTVLTPLIDAGDTFGNALFEVPOQVHRPSSISEMSQSLNDLTHSIPDLAPYQDEKS 2019  
 QY 1392 LDA-PSWPGYVRLSMSIMEQLKTLR-----YNFLPEALDF-- 1426  
 DB 2020 TDLALP-----EVALVEPILAVLQLLCENHNSLONPLRKQSDRTNHNULVSTLSFLD 2072  
 QY 1427 ---VGHQ-----ERTLQ-----CLNAVRTVQSLACEPADHTVGFILQ 1462  
 DB 2073 TVCGSTKSGIVGEIGEENHFSILITQTLATLTERCQCPCHENQNTAMQENGLNIISLV 2132  
 QY 1463 LSNFMKEMPHILPOLMEDIQVNLGYLQACTSLHSKMLQHYLQ--KNGDGLPSAAQR 1521  
 DB 2133 LNEIKPLADDM-ELALEISQASKLALAMESHBDEANRVLRNANNSGGKQVLVA 2191  
 QY 1522 VQPPSAASAAPSSSKQPAD-----TEASEQALHTVQYGLKLSIKTLALRH 1571  
 DB 2192 IKQAVENTNNHMLKSIISDLFRQAEDDLKKSQGPITVNTVTLPEINVDASIGVSI-H 2250  
 QY 1572 FTFDVCOILIDQ-----SLDLAEVNF-LPALSFTPTTFDSEVAPSGTILATVNVALN 1623  
 DB 2251 TEKNISSSLDDKFNDDDIPSDVPREVGNHNYIILAHQALAHIDGEIE-----I 2296  
 QY 1624 MLGSLDKKKEPLQAVGLSTQAGT-----RTLKSLMLTMENCFTYL 1665  
 DB 2297 WLQSDKQDLTRE-ALNTYKERTAOIEIVRRDRIERVVPINDICSTL 2346

RESULT 23



A40743  
 Ipi receptor, XIPR - African clawed frog  
 C/Species: Xenopus laevis (African clawed frog)  
 C/Date: 03-Mar-1994 #sequence\_revision 18-Nov-1994 #text\_change 09-Jul-2004  
 C/Accession: A40743  
 R/Kune, S.; Muto, A.; Aruga, J.; Nakagawa, T.; Michikawa, T.; Furutachi, T.; Nakade, S.;  
 Cell 73, 555-570, 1993  
 A/Title: The Xenopus Ipi receptor: structure, function, and localization in oocytes and  
 A/Reference number: A40743; M01D:93258819; PMID:8387895  
 A/Accession: A40743  
 A/Status: preliminary; not compared with conceptual translation  
 A/Molecule type: mRNA  
 A/Residues: 1-2693 (KUM-  
 A/Cross-references: UNIPROT:Q91908; GB:D14400; NID:g464219; PID:BA003304.1; PID:g464220  
 A/Experimental source: oocytes and eggs  
 A/Note: sequence extracted from NCBI backbone (NCBI:131713)  
 C/Superfamily: inositol-trisphosphate receptor

Query Match 1.64; Score 144; DB 2; Length 2693;  
 Best Local Similarity 17.54; Pred. No 1.2; 576; Indels 776; Gaps 95;  
 Matches 345; Conservative 270; Mismatches 76; Indels 776; Gaps 95;

QY 2 IRKSTKTSVLSFCRSRSLWTLLGRSALRELSQIEALNKWRRLLEGLSYKPPSPSS 61  
 DB 595 ITAAIDTFVSLVRKRE-----PRFLDYLSDLCVSMNK-----SIFVT 633  
 QY 62 AEKV-----KANKDVASPLKELGRISKF-----LGIDEEQSVOQLQCYLQ 102  
 DB 634 QELICRAVLPANADI---LIEKTLVLSREFEGVAGSESLAGEDEER-VWLFWRASN 689  
 QY 103 EDVRGRDVSQVTVLDEROSQALILKIDYV-YE-----ERTC----- 139  
 DB 690 KEIRSR--SVRELAQDAKEQKEDQVLYRYQALNFAKMCIDROYLANEISGQD 747  
 QY 140 -ILRCVLLHLLTYPODERHPYRVEYAD-----VDLKEKELVS--KYRQOFEELYTEA 189  
 DB 748 LILRCM-----SDENLPFDLRASFCLMLHMHVDRDPOEQLTPVXYARLMSIPEIA 800  
 QY 190 -PWETHGNL---MTER--QVSRFPVOCLEB-----QSMLEELI-----FLY 225  
 DB 801 IDVDSDSGTSRDIDIKERFAQTMFEVEYELRDVVGQRPADKEKNKLTFFVNLARNLIY 860  
 QY 226 YAFEMAPSDLLVLTKEF-----KEQFGSRQTNRLHVDETMD 263  
 DB 861 FGYVNF--SDLLRLTKLLAILVDCMILIANPSSKGLKGBSKSGSVNWSIHGVGELMT 918  
 QY 264 PFVDRIQYF--SALILVEG-----MDIESLHKCALDRR--ELHOPADGLICQDMD 312  
 DB 919 QVLLRGSGFLPMTPMARBEGLIKQREPEKEDLVMDTKLIEILQFILNVRLDYRISC 978  
 QY 313 LMLTPGDIPIHAPVLLAMALLRLHTLANPEETSSVVRKIGTALQLNV-----FOYLTREL 366  
 DB 979 LLCIF-----KSEFDESNAQSVESGTEATVVPRTGLDFEHEIEQA 1018  
 QY 367 QSLASGNDCTSTPAICWCVGLSFLVLSLEHLTLGQODIDPACGVLDAPSLPELFWG 426  
 DB 1019 EGFPGSGEE-----NTPLDLDGGR----- 1039  
 QY 427 TEPTSGILILDSVCGMFRHLSPLOL-----RALVSGK--STAKK 467  
 DB 1040 -----TLFRLVLLHLLTMHDYRPLVSGALHLRLFRHFSQROEVLQAFQVQLVLTSGQVNDYKQ 1095  
 QY 468 VYFGLDMSYF--NELYKAK--PHUVISHEDGTLRRQTP----- 503  
 DB 1096 IKODLQRLSRIKSELSLMVYKSGSEVTAQAQAGAKGETPGAKKSESTSYNRYVVK 1155  
 QY 504 KLLYRGGGQNLRIPOGTGVQVMDLDRAYLVREMEYSMTLFCBIEM--LHVSTADV 562  
 DB 1156 EILRL---SKLCVQENTTGRNRKQOORLRNKAIS-----VLEELQIPREKEDT 1206  
 QY 563 ICHQGVKPIIDLNVKIVST-----DLSIADCLLPITSRIYML---QR 603  
 DB 1207 -----RMQELIMKLAHEFLQNFCAQNGQNALHNGHINLFLTPGLLEAVNTQHIHNNFOL 1261

QY 604 LTTVISPVDVIVASVNCVCLTVLAARNPAKWTDLRHTGFLPFY-----AHPVSSLSOM 656  
 DB 1262 CSSEINERVOHFACHE-----TIGRNVQYIKFLQTVYKAGRGYIKKQODI 1307  
 QY 657 ISAEAGNAG-----GYGNL--LNNSEPOGB-----YGVYIAFLRLITTLVK 697  
 DB 1308 VMLELVNNGSEDLVFPYNDRASFOVLQVMRSEBERDENSLMYHHLVBLAVCTEGKN 1367  
 QY 698 QLGSTOSQGLVP-----CWFV-----LKMLDPSYHKRY 727  
 DB 1368 VYTEIKNSLPLPDIDYRVVTHEDCVPEVXIAYINFLNHCYVDTEVEMKEIYTSNMWK- 1426  
 QY 728 NSGVNRQICGLLELILHALNCHERDHSHTPSLQFCISLAVTEAGOVINIMGI 787  
 DB 1427 -----LEENFLVDICRVGNST--DRKAD-----YLERVYTELVMSIVN--- 1465  
 QY 788 GVDITDMAAOPRSDGAEQOGOLIKTVKLAFSVYNNVIRLKPSPNVVSPLEQALSQ 847  
 DB 1466 -----TFSSPSFDQS-----TLQASKLAV-----VHQTRQP--VFVQLQAFIR 1504  
 QY 848 HGAHGNLLIAYLAKYTH-----KHDPALPRLAIQLKRLA---TYA--PMSVYACGN-- 896  
 DB 1505 -----YHCNMLPSPQKASYESCIRVLSDVAKSRITAIPIVDLDSQVNNLP 1549  
 QY 897 -----DAAIRDAFLT-----RLQ---SKIEPMRIKVMLEF 925  
 DB 1550 LKSHNIVQKTAMGWRMSVRNARBDVLTASRDYRNIIERLQDIVSLEL--RLRPLVQAE 1608  
 QY 926 LTFVAVER--OPGLIEFLNLEVDGSDGSEFLGMSG--LHAVLELDSQOQDRWYC 980  
 DB 1609 LSVLVVDVLRPEL-----PEPNTDARKKCSGGQICGLIHNTKLLLENEEK----- 1656  
 QY 981 PLLHRAALIFLHLMODR---RDSAMVLRKPKFW-----ENLTS---PLFPT 1024  
 DB 1657 --LCIKVLTREMMAKDRFGDKGEALRYLVNRYGNIIPAGRRSLTSFGNGPISGG 1714  
 QY 1025 LSPPS-----ETSEPSILETCALIMKIICLEIYVYVKSLSQSLDQTLKKFSIEK 1074  
 DB 1715 SSGKSGGGSIGSSLSRSEMSLSD-----VQCHLDKQGSGLDVI--DLINMATSIR 1763  
 QY 1075 RFAVWSGVVSLAVHVAETEGSSCTSLLEYQMLVSAMRMLLIATYADIMHLDTSVVR 1134  
 DB 1764 VFH-----ESILALALEGGNT-----IQHSFCRLTEDKKE 1798  
 QY 1135 QLFLVDLDGTAL--LVPAVNCYLRGSMKCTLLILRLQWRRELSVDEILGPLETEI 1191  
 DB 1799 KPFKVFEDRKKVAQOEIKATVNTSILSGSKED---EEOGERETPHHQVAREPSGOI 1853  
 QY 1192 LBGVLADQOLMEKTA--KVF-----SAFTV 1217  
 DB 1854 TE---EAKQELIRASVATKKAAYSFRREADPDHFSIGEGVMAVAEKGRDELSAVITI 1910  
 QY 1218 LQ-----MKEMKVSIDIPOYSQVLN-----VCETLQ----- 1243  
 DB 1911 MOPILRFQLCCENHRDLQNFRCQNNKNYLVCTTLOPFDICIGSTTGILGLYI 1970  
 QY 1244 -BEVIALFDQTRSLALGSAETEDKSMETDCSRSHRRDQDVCVL-----GLHLAKEL 1297  
 DB 1971 NEKVALINQLESL-----TEYCOGCPHENON---CIATHESNGIDITIAL 2014  
 QY 1298 CEVDEGDSMLQVTRRLPILPTLLTLEVSIRKQONHFEATLHLLTLARQOGATV 1357  
 DB 2015 TLND-----INPLGRKRMDLVLELKN---ASKLLAIMESH-- 2049  
 QY 1358 AGAGITQICLPILSYQSLSTNGTAOTPSASRSLDAPSPGVYRLSMSIMEQLKTLRY 1417  
 DB 2050 -----DSENAER-----ILY 2059  
 QY 1418 NPLPEALDFVGVHOERTLQCLNAVTVQSLACLSEADHTVGLTLQSLNFKMKNHFLPOL 1477  
 DB 2060 NMRPEKL--VEVIRKAYLQ-----GEVEFEDGESG-----EDYASPR- 2095

```

QY 1478 MRDIQVNLGYLCAQCTSLHSRKMLOHYLQNKNGDGLPSAFAVRVORPPSAASAAPSSSK 1537
DB 2096 -----NVG-----NHYTLAQHARHNE-----LQHMURP-----GV 2123
QY 1538 QPAADTEASEQOALHTVQYGLKILSKTALRHFTPDVCCILLDOS 1584
DB 2124 QTGEGDEALBFYAKHTAQIEIVR-LDRIMEQIVPVPVSGEFLTKES 2169

RESULT 25
S45340
FKBP-rapamycin-associated protein (FRAP) - human
C/Species: Homo sapiens (man)
C/Date: 10-Dec-1994 #sequence_revision 10-Nov-1995 #text_change 09-Jul-2004
C/Accession: S45340
R/Brown, E.U.; Alber, M.W.; Shin, T.B.; Ichikawa, K.; Keith, C.T.; Lane, W.S.; Schreiber
Nature 369, 756-758, 1994
A/Title: A mammalian protein targeted by G1-arresting rapamycin-receptor complex.
A/Reference number: S45340; MUID:94277209; PMID:8008069
A/Accession: S45340
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-2549 <BRO>
A/Cross-references: UNIPROT:P42345; GB:I34075; NID:G508481; PIDN:AAA58486.1; PID:G508482
C/Genetics:
A/Gene: GDB:FRAP1; FRAP; RAFT1
A/Cross-references: GDB:597698; OMIM:601231
A/Map position: 1p36-1p36
C/Superfamily: yeast TOR2 protein

Query Match 1.6%; Score 143; DB 2; Length 2549;
Best Local Similarity 17.6%; Pred. No. 1.3;
Matches 262; Conservative 246; Mismatches 559; Indels 420; Gaps 65;

QY 237 LVTTMKKEGFGSRQNRHLVDETMDP-----VDRIGFSAIIVEGMDIE 284
DB 217 LITLTQREPKEMOKPCQVYRHTFEAEKGFDETLAKERGMNRDRI-HGALLIINELVRIS 275
QY 285 SLHKCALDDBRELHOFADGLI-----QDMDCMLTFGDI PHHAPVLLAWALLRHITLNP 340
DB 276 SMEGERL--REHEBEITQOOLVHDYCKD---LMGFTPRHITPTTSE---QAVQPO 325
QY 341 ETSSVVRKIGSTAIQLNVFOYLTRELLQSLASGNDCTSTACMCVYGLSFVLTSLHT 400
DB 326 QSNALVGLGYSSHQ-----GLMGF----- 345
QY 401 LGNQODIIDRACVLDPSLPELFWGTEPTSGIILDSVC-----GMPPHLS----- 449
DB 346 -----GTPSPAKSTLVESRCRDLMEKFDQVCCQWVLC 380
QY 450 -----PLQLLRATVSGSTAKVYSFLDKMSFYNELYKHKPHDIVISHEGDTLWR 499
DB 381 RNSKNSLIQWITLNLPRLAAPRSA-----FTD--TOYLODTMNHVLSYKKEKERTAA 434
QY 500 KQTPKLIYFGGQNTLRIPOGTVGQVMD--DRAYLVMEYSYS-----SWTLFTCEIE 551
DB 435 QALGLLSVAVRSEFKVLPFR-----VLDIRAALPEKDPAHKQKAMQDAIVFTC-IS 487
QY 552 MLHVVSTADVIOHCQRKPIIDLNVKIVSTDSLADCLLPISRYMLLQRLTYYISPP 611
DB 488 MLARANGPG--IQ-----QDIKELLEPMLAVGLSPA-----LTAIVYDLSRQIPQ--K 532
QY 612 VDVIASCVNCLTVLAARNPAKWTDLRHNTGFLPFVAHAPVSLSQMISAEGMAGAGNLL 671
DB 533 KDIQDGLKMLSLVLMHKP-----LRHPMPKGLAHQLASPLTLTPRESDVG----- 580
QY 672 MNSEPOQGEYGTIAFLRLITTLVKQO-----LSTQSGVLVPCVMFVLKEM 718
DB 581 -----SITLALRTLGSEFEHSHLTQFVRHCADHFNSEHKIRMEARCTSRLL 630
QY 719 LPSYHKRVNASHVRQIGLILIELHAL-----NLCHEIDLASSHT 761
DB 631 TBSIHLSGHAHVASQTAVQVADVLSKLVLGVTDPDDPIRYCVLASLDERFDALHAQ 690

```

```

QY 762 PSIQFLCISLAYTEAGQYINIMIGVDPTIDWMAQPR-----SDGAEQ 808
DB 691 ENLOAL-----FVALNDQVEFIRELAICTVGSLSMNPFAFVAPFLKMLIQILTELEHS 744
QY 809 GOGOLLIKWYKLAFTVNNVIRLKPPSNVSPLEQAL-----SQHAGNNILAVL 859
DB 745 GIGRIKESQAKMGLHVSNAFRLRP--YMEFLKLLIKLQDPDDPENGVINVLATI 802
QY 860 AKYIYKHDPALPRALIQLLKRLATVAPMSVYACIGNDAALRD-----AFLTRLSKIE 914
DB 803 GE-----LAQVSGEMRKVWDELFIINDMLQDSLSLAKRQVALW 842
QY 915 DMKIKMILEFLTVAVETQGLIELFLNLEKVDGSGSKSPSISGMMSCHLAVLELDSQ 974
DB 843 TLQALVASTGYVPEPRKYPTLLEVLNLTAKTEQONGTREAIRVGLLGAL----- 894
QY 975 QDRYWCPEPLHRAAIFALHMODRDSAMLVLTQPKFWMENLTSPFGTLSPSETSEP 1034
DB 895 -----DPYKHKVNLGMD---QSRDASAV-----SLSSKSSQDS 926
QY 1035 SILETCALIMKI--ICLEIYVYVKGSLDOSLQDTLKKFSIEKRFAYMSGVYSLAVHAE 1092
DB 927 SDYSTSEMVLVNMGNLPLDEFYPAVSWV-----ALMRIFRDQSLSHHTWVQATFIFK 980
QY 1093 TESSGCTSLIEYQMLVSAKRMILLITTTADIMHLTDSVVRQLPLDV----- 1140
DB 981 SLGKLCVQPLPOVM-----PTFLNVIYRCDGAIREFLFQQLMVLVSFVSKHIR 1028
QY 1141 --LDGTRKALLVPAVSNCLRLGSMKCTLLILRLQKREIGSVDELIGPTEILBEVLQA 1198
DB 1029 PVMDELVTLMREHWNT-----SIGSTIIL-LIEQIVVALGGEFKY--LPQIIPMLAV 1081
QY 1199 DQQLMEKTKAVPSAFTVLOKMEKVSIDIPOYSQVLNVCTLOEVALPDQTRHSJ- 1257
DB 1082 --FMHNSPGRIVS--IKLILAIQLFGANLIDYLLHL-----PRIYKLFAPAEPAP 1130
QY 1258 ---ALGSATBCKDSME--TDCCSRSHRQDRGVCTGLHAKICEVDEDDGSMV--QV 1310
DB 1131 SRKALLETVDRLTESLDFTVYASRIH-----PIYRILDQSBELRSTMTDLSSLVFO 1184
QY 1311 TRRLPLPTLITLLEVSLRKNLH-----FT-----EATLHLTLTARTOQAT 1355
DB 1185 GKRYQFIIPMNKVLVRRHINRQYDVLCRIYKGYTLADBEEDPLIYGRMLRSQDGA 1244
QY 1356 AVAAGITQISICPLLSVQLSTNGTAQTPSRSKSLDAPSW--PGYTRLSMSLM--EQLK 1413
DB 1245 LASGPVETG---PMKQLHVTINLQKAWGAARVSKD--DWLEWLRRLSLLELDKSSSP 1298
QY 1414 TLR-----YNFLPEALD-----FVGV-----HOERTQCLNAVARTVQSLACLEEA 1453
DB 1299 SLRSCMALAQAIN--PMARDLFMAAFVSCWSLINEDQDELIRSLFALTSDIA---EV 1353
QY 1454 DHTVGFITQLSNPKMEWHEHLPLQMRDIO--VNLG---YLQACTSLHSRKMLOHYLON 1508
DB 1354 TOT--LNLIAEFM--EHSKGPPLRPDRDNGIVILGBRAKCAVAKALHYKEI--EFGG 1407
QY 1509 KNGDGLPSAVA--QVRQPPSAASAAPSSSKQ--PAADTEASEQOALH 1552
DB 1408 PTPALIESLISINNKLOPEEAAGVLEYAMKHFGELEIQTATYERKIH 1454

RESULT 26
D88450
protein P21H1.2 [imported] - Caenorhabditis elegans
C/Species: Caenorhabditis elegans
C/Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
C/Accession: D88450
R/Anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2016, 1998
A/Title: Genome sequence of the nematode C. elegans: a platform for investigating biology
A/Reference number: AY5000; MUID:99069613; PMID:9851916
A/Note: see websites genome.mucsl.edu/gsc/C_elegans/ and www_sanger.ac.uk/Projects/C_eleg

```

A>Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and  
 A:Accession: D88450  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Releases: 1-2700 <STO>  
 A:Cross-references: GB:chr\_III; PIDN:AA19433.1; PID:g506795; GSPDB:GN00021; CESP:F21H11  
 C:Genetic:  
 A:Gene: F21H11.2  
 A:Map position: 3

Query Match 1.6%; Score 142.5; DB 2; Length 2700;  
 Best Local Similarity 17.8%; Pred. No. 1.6;  
 Matches 334; Conservative 261; Mismatches 661; Indels 619; Gaps 92;

```

Qy 145 LHLT-----YFDEBRHYREYADCYDKEKEVSKRQGFEL- 184
Dy 22 LHIYTEERPLEKMLNKTLORGDLYFDMLCHTGLSEICLPEPLIKVLEWYEKDESLC 81
Qy 185 -----YKTBAPWETHGNLWTERQVSRMFVQCLREQSMLEIIFLYAYAFEMAPSDL- 237
Dy 82 LSLMSPILATPELRKLKHLKLANV-----YLFCL-----VLIEI--LPQVEHLRQCDPLI 130
Qy 238 --VLTWKEK-----QGFGRQTRHLVDETMDFVDRIG--YFS----- 273
Dy 131 KKVLEICFKNQVREPSAVGINKTNHLVAETGEVGVLSSTYETHIHRIFMTHITELK 190
Qy 274 -----ALLVEGDISLHKCALDREHLQFODGL-ICQDMDCMLTF--GDI 320
Dy 191 KDVSTAAQOIVAIIMSKPLRINSQVED-----FENGKFLDDGSLLEVKQDGV 243
Qy 321 PNHAPVLLAMALHRTLNPEERSVVRKIGTAIOANFOYLTRLQSLASGNDCTTS- 379
Dy 244 KHAWGLLVELL-----PVAQIKR-----ETNI--PALISLVQKLYTTNDMSKK 289
Qy 380 -----TACMCV-----YGLSPV--LTSLEHLTGNQODIIDTACEVLADPSLEL 423
Dy 290 QHKLAPFLITCLCVSQKHFFLANWQFLNSCSHLKXKDPQVARVLE-----SLYRL 344
Qy 424 FW-----GTEPTSLGIIIDSVCG--MFP--HLISPL--LQILRALVSGSKTKKYS 470
Dy 345 LMVYMRNADGNATRSRLDSICSLPRKGNRYVPRDAPLANIFVKIHLISQKIDFA 404
Qy 471 FLD-----KMSFYNELYKGRPHDVISHEDGILMRQTPKLIYPLGGQNTLRIP 518
Dy 405 FKEIIFDLICVNNRQGRSLYERKNVNGIRALMVAIDGLOQKDDPRAMKSGPSA--- 460
Qy 519 QGTGQVMDLRAYLVREYSSWTLFTCEIEMLLHVVSTADVIQHCORVKKPIIDLVK 578
Dy 461 -GTVHKT--KRRQYITR-----PLTNEISKSIDQPYPCRKADSIIR 502
Qy 579 VISIDLSIADCLPTTSRIYMLQRLTVISPRVDVIASCNCLTVLAARNPAKWTDLR 638
Dy 503 LLDIOIGPLMMSSIQNGKEPDELISGDAKPKDLFRPTCI-----AALPRLLPDWMS 555
Qy 639 HTGFLPFAHNVSSISQMSABGNMAGVGNLMMSEOPGEYGTIAFLIITLVGQ 698
Dy 556 HVDLIDL--LTRLTVHDLBELRMSCGTLTQTIIGEPDMREBOFISHISIOSHI-- 608
Qy 699 LGSTOSQGLPCVM--FVLKEMLPYHKMR--YNSHGVEQIGCLILEIHAHILNCHETD 755
Dy 609 -----YQFFQIILDDSLRLMLQSLTWTKTAITAKKRE-----AKKIIPSPCHSSH 657
Qy 756 LHSHTPELQPLCISLAYTEAGQTVINIGIGVDTIMVMAAQRSPGABGQGGQLLI 815
Dy 658 VTSNSTHSTHISVSPNTVTSASISLSSSSPHQTPS--LCSLPRESSLSHS-----I 708
Qy 816 KTVKLAFAVTVNNVIRLKRPSSNVSPLEQALSQHGKNNLIALVLAK-----YIYKHDP 870
Dy 709 PTAMTSLSTTTTPHISQAPSTSTPI-----GSGNSTLSVLHQMEGLAIVYLCQTRS 761
Qy 871 LPR-LAIQLLKLATVAPMSVYACLDAAAIRDAFLTRLQSK-IEDMRIKYMI---LEF 925
Dy 762 NPRAIAVILKEIQIYIDLLGIEIMIDTVIVIDLQATPYVVKYIEHPIKERMSWMNDF 821

```

```

Qy 926 LTV-----AVETPGLELEFLNLEVKDSDGSKESLGMSCL---HAVLEIDSOODR 977
Dy 822 SSVCDDKSTIETDNTLV-----NSDRGENEYFQNDPMGCALSGAEIHLITK----- 868
Qy 978 YWCPRLHRAAIAELHALMODRDSAMLVLTKEKFEWENTLSPFLGTLSP--PSETSEPS 1035
Dy 869 --CP-----SAVA---AAM-----PILFSRLNA-VSGYVDPMNPQNSRSS 903
Qy 1036 ILETCALIMKICLETIYVVGKSLDQSLKDTLKKFSLIEKRPAYSGYVKSIAVVALEEG 1095
Dy 904 LL-----RGS-----XSKG 912
Qy 1096 SSTSLLEYQW---LVSAMRLIIATTHADIMHLTDSVVRQLFLVDGTAKALLVP 1151
Dy 913 TS--SILGEQLGQACALMOKYILMCA-----LAP 942
Qy 1152 ASVNCRLRGSMTCTILLRLQWRRELSVDEILGPTLEIGVLQADQOLMEKTKAVF 1211
Dy 943 APYNLS-----QRSFPTSMQGPV-----DVF 965
Qy 1212 SAFTTVQMKMKVSDIPIQISQVLANCETIQOEVIALFDQTRSLALG----- 1260
Dy 966 RSVSASIRSSRTPVPN--SISQILSKVCMILRWENTL--DIRDSVVLGVGSINPLAFDM 1020
Qy 1261 -----SATPDKDSMETDSCRSRHRQD-----GVCVGLHIAKEICEVED 1303
Dy 1021 MLEELKNSGILREATEK--AETN-----LRRKKKDLRLQITIVIEVALRGLLHSA 1074
Qy 1304 GDS-WLQVTRRLPIPLTLTLE-----VSLRMKONLHFEATVHLTLTLARTQOGATA 1356
Dy 1075 GSSDFILHPRVVDIDSMRVLESBDHRODITVYRKLHFAK-LIHILID----- 1123
Qy 1357 VAGAGTOSICLPLLS-----VYOLSTNGTAQTPSASRSKSIDABSMGVYRLS 1404
Dy 1124 -----STPHLSHHTLPSDERRHKLPLYLFIMGCSRAIAADRKFKDKSVGSYVEQKS 1173
Qy 1405 MSLMEOLL-----KTL-----RYNPLBEALDPVGHQERTLOCLANAVTQSLACL 1450
Dy 1174 VLANSRLICCGPIREPAKSGIEGDYLGWLEKLI--ISTNPTWQA----- 1216
Qy 1451 BEADHTVGFILQI--SNFMKEW-----HFHLPLQI-----MRDIO--VNLGY 1487
Dy 1217 -EVEBMLAMMELMESGVLDMVMQCTQPTVACGCFRALVRFSSRDPCEFSLFV 1275
Qy 1488 LCOACTSL-----LHSRKM-L-QHYLQNKGGDGLPSAVQVORPPSASAAASSSKOP 1539
Dy 1276 LCOGMLAVNSVTDICALHMBILIRKQPLETSN-----IHATSPAQCVTEIVQWRN 1324
Qy 1540 AADTBASRQALHTVQGLKILSKTLAALHFTPRVCOQLLDGSLDLAEYNFLFALSFT 1599
Dy 1325 QADVLTNTVNGCH-----ILPIEGB--DVCTRIAN-----STPHLTVTIFS 1363
Qy 1600 TPTF--DSEVAPSEGTLLATVVALNMLGELDKK-----KEPLTO----- 1637
Dy 1364 EVSYRLTESGNSNSQILLALQPMISNLELVDOUVVEBAAGPRGMSGEATQVLANNLL 1423
Qy 1638 --AVGLSTQABGRTLSLMLFTWNCIFYLLISQA-----MRYLRDAVHPRDK- 1684
Dy 1424 YLVTVLSDNDEKELAEBSLIPFTYRIC--VMSQVVGTRISSILIEWL--STVHDSKIT 1479
Qy 1685 -----QMKDELSELSTLSSLSR--YFRGAGPSRPATGVLPSPQ--GKSTSSLSKA 1732
Dy 1480 LDRSEIIPPYRWKDE-NSERVANGSIARDSLREBEPKGVRLPMPAYGHHYSPLOSQF 1538
Qy 1733 SPSEQPEILOVQAF 1747
Dy 1539 LP-----PYVQPVQFF 1549

```

RESULT 27  
 A54837  
 rapamyclin/FKBP12 target 1 - rat



Query Match 1.6%; Score 141.5; DB 2; Length 2114;  
 Best Local Similarity 17.9%; Pred No. 1.2; Mismatches 646; Indels 561; Gaps 77;  
 Matches 320; Conservative 262; Mismatch 646; Indels 561; Gaps 77;

47 LLEGLSYKPPSPSSAEKVKANKDVASPLKELGRI--SKPLGDEOSVQ-LLOCYQOE 103  
 195 ILKLLQSSNPVSSQNA-----ASLLARL-RIPTSSISKVESAGVQLVLLGEE 244  
 104 DYRGTRDSVKTVLQDEROSQALIKLADYYEERFCIRCV--LHLTYFODERHPYVE 161  
 245 NSVFRAVSVAAL-----EAITSK-----SEALTVARDLDGHLI----- 281  
 162 YADCVDKLEKELVSKYRQOFEELYKTEAPWTEHGNLMTEROVSWFQCLREOSMLEI 221  
 282 -----SAVVASKESSVE-----ETERVLOSCTGOTALANLCGMSG 317  
 222 IFLYAYVEMAPSDLLVLTKEKQGFSGRQTRHVLVD--ETMDPFVDRIGVFSLILVE 279  
 318 LIVVLGSLSPRLTEPIADILGALAYALRKQFQSGDTRFADF-----TLTE 366  
 280 GM-----DIESLHKCALDRRELHQPADGLICQMD-----CMLTFQDIPHPAP 325  
 367 GILVKLKPRTQILHERILFAMESLFGNVDSLKLNVDAKRYLVCLTITATDGPBRM 426  
 326 VLLAMALLRH-----TLNPEETSSVVR 347  
 427 ITCSNLCKHGDWDALGREGIOQLIPYLGSSQHOBLSEFLAITDNVEBSRAVVT 486  
 348 KIGGTALQANFOYLTRLLQSLASGND-----CTTSTACMCV--YGLLSF 391  
 487 SAGG-----IPLLQILETGVSQKAKODAVRVLNLCHEBEIRLCYERKAGALPA 536  
 392 VLTSLHLLTNGQDIIITACEVL--ADPSLPE--LPWGTEPISGIGIILDSYCGMP 445  
 537 LLGLKNGKSPKQSSANTLLKLTADPSVIEQVAFGLGAPSKTHLI--RVLG-- 591  
 446 HLLSPLOQLBALVSGSTAKKVYSLDKMSFYNLYHKHPDVI-----SHED--GTL 497  
 592 HVLAS-ALBEFVYTGSAANGRLSVORLASNEMKENASVLAADFPSSKIDCGELG 650  
 498 WRROTPLKLYPGGQTNLRIPQGTIVGOVMDRAVLVMEYSWTLFTCEIEMLLHV 557  
 651 FDEBDNPTCKLLSGNT-----HAV 669  
 558 STADVIQGCQKVPITIDVHKVISTDLSIADCLPITSRIYMLQRLTTVISPPDVYAS 617  
 670 AT-----QLAHALGS-----LSNPTKKTATKLS--GPEYEVKTP 703  
 618 CVNCLTVLAARNPAKVTDLNHTGFLPFVAHPVSSLSQMS-----AEGNNAAGVGNL-- 670  
 704 LIKS-----AKTNP-----TSTENPMSTLANLSDPNVAALDNDVVSALTR 747  
 671 LANSBOPOGEYGVTIAPLRL-----ITLVKSGQLSGTOSQGLVPCWVFVLEKMLPSYHKW 725  
 748 VLREGTLQCKGNASHALHQLKHQVSDVFG----- 779  
 726 RYNSHGVRQIGCLILELIHALINLCHETDLHSHTPSLQFLCTISLAYTAAGQVINIM 785  
 780 -----NEOCRFAVSELI-----DLNATNDLANS-----AF----- 804  
 786 GIGVDTIMWMAAQRSDGAEQGGQGLIKTVKLAFSVTNNVRL--KPPSNVSPLEQ 843  
 805 -----IDVLEVL-----LLAKAKYGANLSHNPSPSAFEBVSNDLSLVRG 844  
 844 ALSQGHAGNNLIVLAKYIYHKDPALPRILAQLKELATVAPWSVYACIGNDAAIRD 903  
 845 LAEGHPLVQDAIELISRFCKTQF--ILGRLLVTSKSISSLANNTINS--SSPEIKVGG 901  
 904 AFLRLQSKIEDMRKVMILEFVIAVETQPGILFLFANLEVKDSDSKSPSLGM----- 959  
 902 AILLVCAKND-----ITLMEAVE--QSGYLLTVLNTLLDMSKNSKASAYGIEIQR 952  
 960 ---WSCHLAVALLEIDSGQODRYWCPLLHRAIAFLHLMQDRDSANLVLRTRKPFEN 1016

DB 953 PRSFTTNSLCIRMDSEKNDV-----TITGTSAMWMLSTIICSSHPNSRLV--MENGLEI 1008  
 1017 LTPPLFGLTSPSPSTSEPSILETCALIMKIIICLIIYVVKSLDQSLKDTLKTSIEKRF 1076  
 1009 IAEMLQNRKSNVQNSSDSESEKMIAMGFLVMSQEPKVSAPATENIIQTLAPMQSOM 1068  
 1077 --AYMSGVKSIAVHVATESSCSTSLLEYQMLVSAKRMILLIINTHADINHLDSVVR 1134  
 1069 IDGYFTAQV--LAALVRRKDKTISEIMNSDIVET--INLVGCEBSDRSLCALAEEL 1123  
 1135 QLPFLVDGTKALLVPSAVNCLRISGM--KCTLLI-LRQMRRELSVDEILGPTLEI 1191  
 1124 SLVQNPYEATILEVLFENERV--RSGSTTKKCIPLVNLKPYADKKGIFVALRLRRI 1180  
 1192 LEGVLADQQLMEK--TKARVSAFTVLQMKEMKVSIPQYSQVLNVNCEVLOEYIA 1248  
 1181 -----ADNDLSKILLIAEAGALDALAKYLSLSPQDSTEI-----TVSEILDS 1222  
 1249 LF-----DQTRSLAGSATKEDKDSMETDCCSRRRRDQDVCVIG-LHL-----AK 1295  
 1223 LFRSPELTRKKTALSSMKQ-----LIGLHLASRSTRYNAR 1259  
 1296 ELCEV-----DEGDQMLQVTRRLPIPLTLTLEVSLMKOMLHFEATLHLTLARTQO 1352  
 1260 VLCELFSEHIRDEBLMKALSPLEMLNTLE-----SERVALTLVKTLMGINPRPD 1314  
 1353 GATVAGAGITQISICPLLSVYQSLTNGTAQTPSARSKSLDAPSPGVYRLSMGLEQL 1412  
 1315 ILTSLEGN-----PLDNIYKI-----LSLDS-----SLESKTSAA 1345  
 1413 KTLARVNLPEALDVGCHQETTLQO-----LNAVTVQSLACLEADHTVGFIIQLSNM 1467  
 1346 RICPLFLFNR-----GRTSTSAACCVSLISLRTGSGTA--IAGMFALDRLDIRKRV 1399  
 1468 KEWHFHLPOLMRDIQVNLGYLQOACTSLHSRKMLQHYLQNKNG-----DGLP-- 1515  
 1400 EVAEER-----DCVNLFGYVASENVLISEAALSCITRWAKNTPRKM 1442  
 1516 -----SAVAQRVQRPSPASADBSKQPADE--ASEQALHTVQYGLIKLISK 1564  
 1443 DLKMGIEIKCISQSPSPSLCSVIADLFRVLTNGVIARSODAIKMQVPLLIILRQ 1502  
 1565 -----TLAALRHRTPVQOILLDQSLDLAYNLPLALFTTPPFDSEVASFGLIATV 1618  
 1503 DLDFQGLQGGQAJANILEKMPVLESKIASSTIIMP--IPLESE--SIWKNAT- 1555  
 1619 NVALNMLGEIDKKKE-----PLTQAVGLSTQABGRTKLSILMPTME 1660  
 1556 TILTSLEMRQFQEBITTKLILAPLVKVGIR-----VRNLQRIALMGL 1601

## RESULT 29

750176  
 probable peptide synthetase (imported) - fission yeast (Schizosaccharomyces pombe)  
 C:Species: Schizosaccharomyces pombe  
 C:Date: 09-Jun-2000 #sequence\_revision 09-Jun-2000 #text\_change 09-Jul-2004  
 C:Accession: T50176  
 C:Badcock, K.; Churchev, C.M.; Wood, V.; Barrell, B.G.; Rajandream, M.A.  
 submitted to the EMBL Data Library, February 2000  
 A:Reference number: Z25044  
 A:Accession: T50176  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-4924 <BAD>  
 A:Cross-references: UNIPROT:Q9P7T1; EMBL:AL138854; PDB: CAB72227.1; GSPDB: GN00066; SPDB:  
 A:Experimental source: strain 972h(-); cosmid c23G3  
 C:Genetics:  
 A:Gene: SPDB:SPAC23G3.02c  
 A:Map position: 1  
 A:Insertions: 2125/1  
 C:Keywords: carrier protein; phosphopantetheine; phosphoprotein  
 F:260-697/Domain: acetate-CoA ligase homology <ACLI>



F:2772-3226/Domain: acetate-CoA ligase homology <ACTL2>  
 F:4405-4474/Domain: acyl carrier protein homology <ACP>  
 F:2206,3288/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

Query Match 1.6%; Score 141; DB 2; Length 4924;  
 Best Local Similarity 19.7%; Pred. No. 5.3;  
 Matches 203; Conservative 137; Mismatches 369; Indels 320; Gaps 49;

```

QY 752 HETDLSHTPSL-----QPCICSLAYTEAGQVYINMGIGVDITDV-MAA 798
DB 707 HSTDSSRNQSLVLAANLHETCEKRFNSVVIS-----RFSIFDGLSDLDIVFVR 760
QY 799 QPRSDAGQGQGGQLLIKTVKLAFTVNNVRLKPPS-----NVVSPLEQALSGHGA 850
DB 761 KIRSLGFPANPS---IVTSKVVFKAISLLIKKENVOSKNNFLNCKTPIQKGLYSF 817
QY 861 HGNMLIAVLAKIYKH-----DPAFLPAIQLLKLAIVAPMSVYACGNDAAIRD 903
DB 818 SNNG-----NLVYHNVFVKIAPSPKVKLAMEKL-----LDHTIILSNGFALDEN 862
QY 904 AFLFR--LQSKIEDMRKIMLEFLTVAVETOPGLIEFLNLEVQGS---DGSKEFSLG 958
DB 863 EGFTRFLEKPPKPLYSYKNCLECIQKAFPTTK- FDEQFLNSGFLDAVITYDSNCVSI 921
QY 959 MMSCLAVLELIDSGQODRYWCPLLRRAIAFLHAWQDRDSAMLVLRTPKRWENTLT 1018
DB 922 VMH--HALYD-----GWSIDIIMQGLFHLH---DR-----LTIVPQPED--- 957
QY 1019 SPLQGLTSPSEBTSPLFETCALIMKICIEIYVVGSLDQSLKDTLKFSIEKRPAY 1078
DB 958 -----YV-----QEL-ESLRNLNKNKNCISF 976
QY 1079 MSGYV-----KSLAVH---VAETEGSGCTSLLEYQ-----MLVAMMLIIIA 1118
DB 977 WKYKLDKFEKFSLSGQRKMGVLEISNISILSVENTCKOQITPLSPFLTANSTVL--- 1033
QY 1119 TTHADIMHLDSSVVRQOLFVDLDTGKALLIVPASVNCRLGSKMCTLLILLRQMKREL 1178
DB 1034 -----SSYIKTN---DFLVGT-----VVGSRVNSL-----L 1056
QY 1179 GSVDEILGLPTEILEGVLQAOQOMKTKAK-VFSAPITVLQMKEMVSDIROYSOVLIN 1237
DB 1057 PNVIYVIGPCWQITPVRKIDDELSEYKULCONLFREISFVLKHSVMAISDE-QEELIVSN 1115
QY 1238 VCETL-----OEVIALFDQTHSLALSGATEDKDMETDCSRHRHQRDGV 1286
DB 1116 LFEISILYQSGIPSVBESFISLHSTDH-----VEQPLIREIKNGKVKF----- 1162
QY 1287 CVLGLHLAKELC-EVDEBDGSMLOVTRRLPIPLTLLTLEVSLSMKONLHTEATLHLL 1345
DB 1163 --LTVGLSSELLNNLNDFFK-----ILNFIYNIESKIQTHASFVNTTIEHNHVE 1211
QY 1346 TLARTQCATVAVAGITOSICPLLSVYQSLSTNGTAQTPSARKRSIDAPSPGVYLSM 1405
DB 1212 SKATL---GFSKREKLVRSCLSKILGNTVLSAVFENLQYGDSDCLANRFSLARKSS 1268
QY 1406 SLMEQLKTLRYNFLPEALDFVGNHVERTLOCLNAVRTVOSLACLEBADHTVGLIQLSN 1465
DB 1269 GIGMLKIPMKSNPTIASL-----CELLVLPTEITLS---ADNBITP-YEVSD 1311
QY 1466 FMKEMHHTLPLQMLMDIQVNLGYLQOACTS--LLHSRKQLQHYLONKRGDGLPSAVNRQV 1523
DB 1312 IGHANL-----DISFOYFPTPMQOALLASSEKNGVEYYNKK----- 1351
QY 1524 RPPSAAAPSSKOPADTEASEQOALHTVQGLIKILSKTLAALR--HTTP-----DV 1576
DB 1352 -----LFTGKSSQEBL-----YLLFTLNLNPLIETCTCFVTRSKYXPY 1391
QY 1577 CQILLDSLDLAEXNPLFALSFTTPTPDSVAPSGTLLATVNVNALMMLGELDKKEPLT 1636
DB 1392 CQVVLNE-----PNFYQVLFPYKGSLSKYNLAELPL--LDSKKYPI- 1431
QY 1637 QAVGLSTQABGTRTLKSLIMPTMENCYLLISQMRYLDRPAVHPRDKQMKQELSELS 1696

```

Db 1432 QIFFL--QGEN-----KNYVFCIHNVLY--DAWAFOIIMDINHILKRNPKG--SQSML 1481

QY 1697 TLSSLSRY 1705

Db 1482 KFIISYLHRY 1490

# RESULT 30

H86502 excludinase ABC subunit A [imported] - Chlamydia pneumoniae (strain J138)

C/Species: Chlamydia pneumoniae, Chlamydia pneumoniae

C/Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 09-Jul-2004

C/Accession: H86502

R/Shirai, M., Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shibata, T.; Ito

Nucleic Acids Res. 28, 2311-2314, 2000

A/Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.

A/Reference number: A86491; MUID:20330349; PMID:10871362

A/Accession: H86502

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 11826 <STO>

A/Cross-References: UNIPROT:Q92985; GB:BA000008; NID:98978469; PIDN:BA098306.1; GSPDB:GN

A/Experimental source: strain J138

C/Genetics:

A/Gene: uvwA

C/Superfamily: Chlamydia trachomatis probable excludinase ABC chain A; ATP-binding case

Query Match 1.6%; Score 140.5; DB 2; Length 1826;

Best Local Similarity 19.0%; Pred. No. 1.1;

Matches 365; Conservative 234; Mismatches 610; Indels 717; Gaps 95;

```

QY 223 FLVYAEEMAPSLLV--LTKMKEGFGSGRQNNRHLDNBTMPFVDRIGYFALLIVEG 280
DB 92 FSHYSHATVSGTIELSHALLFTLLEG---QARDPKTEVLDLSKEKYLSTIMELSEG 147
QY 281 MDIESLHKCALDDRREIHQFADQDLICQMDCLMTFFGDIPIHNAPYLLAWALRHITLMP 340
DB 148 VQSIILAPLRKIDIAIHVEYAQG-----FTKYRCNGTIIPIYSFLTSGI-PE 194
QY 341 ETSVSVRAKIGTAL--QANVFOYTLRLQSLASGNDCTTSTACMCVYGLSLVLTSL 398
DB 195 DCS--VDIVDTLIKENNIRARLKSILFTALFEGEHC-----VLDBEEL 238
QY 399 HTGNGOODIDTACEVLADPSLPELFWGTEPTSGILIIIDSV-----GMPPHLSPIIL 452
DB 239 MTSSTGQIDD---VTYRPLRQGLF---SPHA-----LESRSLOQSGSIFISINDPL 286
QY 453 QILRALVSGKSTAKKYVSFLDKMS--FYNELYKHKPHDVISHEDGTLWRQTPKL--LYP 508
DB 287 -----IDENLSIKENCCSPAGNCSSYLHTIYQALA-DALNFWLETPMKOLSP EIQNIF- 339
QY 509 LGGQTNLRIPQGVQVM----- 526
DB 340 LRKKNMLVLEVRLEFDTLGKRNLTYYVWVGVLNDIGDKVRYTTPKPSRYLSKGNASHCSL 399
QY 527 -----LDRAYLVWME-----YSYSSWTLFTCEIEMLHVSSTADVIQHC-ORVKP 571
DB 400 CKSTGIGDVAVATWVGKPTFERQOMSLNNHVFSPKVS--PSLSIQIQLQGLKRLSE 457
QY 572 IIDLVHAKVISTDLSIDC-----LIPITSRYMLQRLTTVISPP-----VD 613
DB 458 LIDLGLGYLTPNPAIATLTSGEQERFAIAKHGLGELFGITIIDBPSIGLHPDTEKLG 517
QY 614 VIASVCNCLTVLAARPAVWTDLRHTGFLPFVAHVSSLSQMSISMGNAAGYG----- 668
DB 518 VI-----KKLRDQGNVTVLVHEBERMISLADRIIDIIGAGAGIFGEGELP 561
QY 669 -----NLNNSQPOGEY-----GVTIAFLRL 690
DB 562 NKGKDEFLNNSSTLTKYVROBELTIPIPESREAPTSWMLTEATTINLKNLSIRLPLADL 621
QY 691 ITTLVKGQUGSTOS---QGLVPCVWFVLEKMLPS--YHKWRVNSHGVRGEGIGCLILELI 744

```



```

Db      622 IG--VTGSGSGSKSLINNTLVPAIESFLKQENPNKHLFEW-----GC-IGSLI 667
Qy      745 HAILNLCHETDLHSHHPSLQPLCTSLAYTBAGTIVINIGVDITDMAAOPRS-- 802
Db      668 HI-----TRDLPGRSORSI-----PLTYIKA-----FDDRRELFASQPRSLR 704
Qy      803 -----DGA--EGOGGOLLKTKVLAFSVTNNVIRLKPENNVSPLE-----Q 843
Db      705 OGLTKAHFENPOQACTICQCGLTMTI-----SDDPTPIPCSECQ 745
Qy      844 ALSQHGA-----HGNMLIAVLAKYIY-----HKGDPLPRLAIQLKRLATV 885
Db      746 GKRAHSEVLLEIYEGSKNIADILDMAYAEKFPISHPKIKHAKLGLRD----- 797
Qy      886 APMSVYACLDAAAIRDAFLTRLOSKIE---DMRIKVMLEFLTVAETQ--PGLTEL 939
Db      798 -----YPLGRPLSTLSCGEIQRLLKLAHELLFASPKQTLVYLDEPTTGLHTHDIQALIEV 852
Qy      940 FLNLEVDKSGSKSPSLGMSCLHVLLELDSQODWYCPRLHRAIA-FLHALMOD 998
Db      853 LLSLTY-----LG-----HTVLVIEN-----MHVVKCDYVLELAGE 885
Qy      999 RRD--SAMLVLRKPKFMENTLSPFLGTLPSPSET-----EPSILETCALIMKIC 1048
Db      886 GGDGLGYTLASCTPRDLQLNTPTAKALAPIESLDIPVKSSEPPSKCDILIK--- 942
Qy      1049 LEIYVYVKSGLDQSL-KDTL-----KPSIEKRPAYWSG-----YVKSIAV 1088
Db      943 -DAYQNNKHLIDLPRNSLIAGPAGSGKSLVFDILVYAGNAYAELEPPYRQQL 1001
Qy      1089 -----HYAETEGSS-----CTSL-LEYQMLVSAW-----RMLIITHTADI--- 1124
Db      1002 KETPLPSVGEVKGSLSPVISVRKCSSNSNYHTIASLGLSNGLELFAILDEPPSPLE 1061
Qy      1125 ---MHLTDSVVRQLFDVLDGTAKALLVPASVCLRLGSMKCTYLILRLQWRE--- 1177
Db      1062 KLSKIHLPSSSIATQSYNDEL---TSLTSP-----IFLGD---LEIFQEKQKEGFI 1109
Qy      1178 -----LGSVDLIGPLTEILEGVLDQDLMEKTKAVFSATVYLQMKEMKVSIDIPO 1230
Db      1110 KLYEGNNLYDDELR-PLNLEIPAIVIOHTKVSFPKSSSLSAISVAPSLSEIWIYISQ 1168
Qy      1231 YSOLVANCETLQEBEVLFPQTRHSLALGSATEKXDMETDCRSRHRQRCVCL- 1289
Db      1169 KKQRDLSTSLGCKDKGRLYPEITHQL-LSS-----DHPERCLUTC 1208
Qy      1290 ---GLHLAKELCEVED-----GDSWLQVTRRL-----PILPILTL 1324
Db      1209 GGRGEILKISLEHKEKIAHTPLFEFSLFPKSYMKVQKLDENASQPL--KLTLT- 1265
Qy      1325 EVSLRMKNLHPTFATLH-----LLTLTARTQO-----GATVAGAAGITQSI 1366
Db      1266 -----KEFLNFCRSSSEFPKNALMLBQDLTESDPLIKPLALATCSPACKGSLANDYA 1319
Qy      1367 -----CLPILSVYQ-----LSTNGTACTPS-----ARKSL- 1392
Db      1320 NYVAINNTSLLDIYQEDATPLESPLNTIGTDTRSIIDLMLNRLTFISKVGLSYITLQOR 1379
Qy      1393 -DASMPGVYRLSNS-----LMEQLKTLRYNPLPEALDFVGHQERTLQCLN 1439
Db      1380 QDTLSDGNYRLHLAKKISINLTNIVYLFEBEFLSLGHQDPLTYQLL---KEVANNMT 1446
Qy      1440 AVRTVQSLACLEADHTV-----GTLQLQSNFM--KEWHFHLPOLM----- 1478
Db      1437 VIATDRSGSLIPHADHAIFLPGSGPQGGFLMDSPTVCPSVDLHANVPQREVCAPL 1496
Qy      1479 -----RDIOVNLGYLQACTSLHSRKMLQHTLQNKNGDGLPSAVALQRPQRP 1527
Db      1497 ISKANHTGSDRTLKVNLSI-----HHIQN-----LKSASLH 1529
Qy      1528 A--ASAPSSSKOPADTEASEQALHTVQYG-----LKLKLSKTLAALRHFTPDVQOI 1579

```

```

Db      1530 ALVAGVSGSGKSTLLBEGFKKQAEHLIAKGTTFSDLVVIDSHPIAS----- 1578
Qy      1580 LLDPSLDAENRFLPALSTFTPTDSEVAPSGFLATVN--VALNMLGEL---DKKPE 1634
Db      1579 --SGRSDISTY-----FD--IAPSLRAFYASTQAKNLNISTWFTNTXQOQ 1622
Qy      1635 LTOAVGSTOAEGRTLKSLMFTMENCYLL-----ISQAMRY----- 1673
Db      1623 CSDQGLGYQ-----WIDRAFYALEKRPPTGSGFRIOPLAGEVLYEGKHFG 1669
Qy      1674 -----LRDPAVPRDKQRMKQELSELSTL-----LSSL-----RYFR 1708
Db      1670 ELHLPLEIVTALRFPFIKIQKP-LKALLDIGLYPLIGQLSLSVSEKTAALKTAAYFLY 1728
Qy      1709 GAPSPP 1714
Db      1729 QTPETP 1734

RESULT 31
D96746
hypochemical protein T9N14.23 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C:Accession: D96746
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federici, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Hutzar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.;
C.A.; Li, J.; Li, Y.; Liu, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maltli, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salberg, S.L.; Schwartz, J.R.; Shim, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: D96746
A:Status: preliminary
A:Molecule type: DNA
A:Releases: 1-2777 <STO>
A:Cross-references: UNIPROT:Q9C7S8; GB:AB005173; NID:g10645381; PIDN:AA621500.1; GSPDB:GR
C:Genetics:
A:Gene: T9N14.23
A:Map position: 1

Query Match 1.68; Score 140; DB 2; Length 2777;
Best Local Similarity 18.7%; Pred. No. 2.4;
Matches 359; Conservative 238; Mismatches 609; Indels 716; Gaps 95;

Qy      96 LLOCYLOEDRGTDSV-----KTVLODERQOALIL-----KIADYVERCIL 141
Db      1110 LVQSVLSEVVDGSKDSCWMLPRLVLLFTQSLSEPLHSGRTTCGLADTSFATLDEI 1169
Qy      142 RCVLHLTYFQDEBHPRVREYADCVDLKEKELVSKYQO--FEBLYKTEAPWETHGMLMT 200
Db      1170 KGLMRSTI-----PDEYGVVNAFSSALICATPESILKNAASVMDIXGSF- 1219
Qy      201 EROYSRMFOCLRBQSMLEIIFYAYF---EMAPSUL---LVLTMRKEQGFSGSRQ 252
Db      1220 -----SFLSLIFLEKNPLGNLSKLSL--DLPMRGLRLT----- 1251
Qy      253 TNRHLVDETMDFPYDRIGYFALIVEG--WDIESLHKCALDDRELHQF--AODGLIQD 309
Db      1252 VSRNLRGTYVDSIDPADHSSTTEKIKSKODIPDISLAFSVLEQTFPVLLNLEIWSMD 1311
Qy      310 MDCL---MLT-----FGDI-----PHHAPVL-----L 328
Db      1312 ISCLPFRRLTELLKLVQSPKSDISIDRLILFMLFOIRSLKYQPH--PVLCOQSEI 1369
Qy      329 AVALLRHTLNEETSSVVRKIGTAIQANVFOYLTR--LLOSLASGANDCTTSTACVCY 386
Db      1370 CLRMRHLFSQISKDLVSGPSADKLKHQVQGVTLSPHVWALLIESPADCTLPRVQNV 1429

```

```

QY 387 GLLSFLVTS-----LEHITLGNQODIIDTACE-----VLADPSLPELFW 425
DB 1430 VPESELLINAGRIGISEIDQHIL-----DLASTCEFLPEESIERKGLRADKSI----- 1480
QY 426 GTEPISGIGIIDSVCNMF-----PHLISPL-----QILRALVSKSTAKKY-----S 470
DB 1481 -----MARKVIVERLLVFRDKEFLCVGSOSYAPILQHOLIQLALKEFISPKLLYLHNS 1535
QY 471 FLDDKSFNELYKHKPHDIVISHEDETLWRQTPKLLYPL-----GGOTNRIQGVIGQVWML 527
DB 1536 MLSK-TYEEEL-----ASPLISFEGDLIAGAFEMLLIYSR--QPA 1573
QY 528 DRAVLVWEYSYSWTLFTCEIEMILHVSTADVIQHOCQVKKPIIDLHVHVISIDLSIA 587
DB 1574 KRRVDFLMELEDKNY-----DSRIIEGVSLA-----CR-----FSTRELA 1611
QY 588 D---CLFETSRIV-----MLORLTTVISPPV-----DVASCVNCLTVLAARPAK 632
DB 1612 SADTCLLVKVGIGIFGNGSQCSCVHQIIVIMSOIVGRTSKDLIHCIN-----QASMSRAK 1667
QY 633 VMTDLRHTGFLPFVAHPVSSLISOMISAEQMANAGVGNLHNS--EQPOGEVGTIAPLRI 691
DB 1668 I-----LFYLVESPLHLSVFGHIFFSMLSKLQGDNA-----LI 1701
QY 692 TTLVKGQLSTQSGQVPCVWFVKEMLPZYHKRKNYSHGVRBOIGCLILELHAIINLC 751
DB 1702 T-----DDQFVMLPPLVLLFLASVPAKLEK-----SC-----SKC 1731
QY 752 HE-TDLHSS-----HPSLQFLCIGSLATTEAGQVYINIGIVDIDIMVMAQRSD 803
DB 1732 LDITSLNISNIIKGFLOMPKFCSCGCTFEKYE-----LILST-----SE 1771
QY 804 GAEGGQCGQLIKTVKLAFSVTNNVIRLKPPSNVVSPLQALSOHAHNNLIVAKYI 863
DB 1772 DIETMFASNLGKAVRM-----FQ 1790
QY 864 YHKIDPALPLAIOQLKRLATVAP-----MSYVACIGNDAAIRDAF--LTRLQSKIE 914
DB 1791 YHFLTESPTKTDLLKVFYMFPHTSAGKEMLDYEIKEDVVKVDDQMFNTNRLVAKVE 1850
QY 915 DMRIKVMLEFLVAVETOPGLIELFNLBYKQSD--GSKESPL-----GMSCL 963
DB 1851 LSRICLPEDSCMHLKQAGCC-----VKSSPEMGSREBLNPLNAFVNTQC- 1902
QY 964 HAVLELIDS-----QQODRYW--CPPLHRAIAIFALHMODRRDSAMLVLRPKF 1013
DB 1903 -----VVERSGYKXGNSEREODKWFCKSLEY-----FIRSLIKF 1941
QY 1014 WENITSPFLGTLSPSETSEPSIILETCALIMKIICLIIYVYVKGSLDOSLQDLKKEISIE 1073
DB 1942 LEGMCEEL-----AHLDSPFLERLMNMLIRYFKDS--KITKILREIFSF 1986
QY 1074 KRFAVMSGYVSLAVHVAETEGSSCTSLBYQMLVSARMMLIATTH--ADIMMLDTSVY 1132
DB 1987 SRGKY-SYHPQDVLSHQFTESISLS-----ISSHTGEVIRPVSSIL 2029
QY 1133 RROFLVDLDTKALLLVPAVNCILRGSMKCTLLILLRLQWKRELSVDEILGPLETEL 1192
DB 2030 K-----LTIIP-VLNSVRVEN--CS----- 2046
QY 1193 EGVLANOQQLMEKTKAVSFATIVLQMKEMKVDIPOYSQVLNVNCEILOEVIALFDQ 1252
DB 2047 -----LEAPEYLSQIILIKLIGVLCKCKDSGIFPKDL 2081
QY 1253 TRHSIALGATEDKDSMETDDCSRRHRDQDVGLGLHAKELCEVDED--GDSMLQ 1309
DB 2082 HFRLLCGVATPPSIIDQS-----YKLMHDIULYGEHTIAND--SEDIYMGNAALK 2130
QY 1310 VTRRLP--LIPTLLTLEVSLRKONIHFE--ATLHLLTLTARTQGA--TAVAGAG 1361
DB 2131 IREBLPEDGSYSYDIEDVLRQSL-FKENLICLDPKRCAGTIVYFPGYRAESDMTYIVDDP 2189
QY 1362 ITGSCICPLISVY-----QLSTNGTAQTPPSASRSKSLDAPSPGV 1400

```

```

DB 2190 ISEK-CSPALERDPPVILAVSVQLSMTYIEVEPSPASGLAVAPRSMSSADL-----GI 2244
QY 1401 YRL-----SMSIMEQLKTRYNPLPE-----ALDFVGHQERTLOCT--N 1439
DB 2245 RKLGETYGLIFVDLEGIFLPHDFGSGSTVNFPSQRLMVRLVFGLESEBDAQIYIRNS 2304
QY 1440 AVRTVQSLAEBA-DHTVGFILQSNFMKEMFHLPOLMRDIOVNLGYL--COACTSL 1496
DB 2305 ILETWGFSSPLADDETKGLIQVVRKSVKLN-----KARHLVENCGLLSWCSFFSML 2360
QY 1497 HSRKMLQHYLQNKNGDG-----LPSAVQVRQVRPSAASAPSSSKQPAATAS 1546
DB 2361 TTK-----PTGDEDSRFVVVLVIDDALSR-----NTDWS 2392
QY 1547 EQQALHTVQGLKLIKISTLAALRHFTPDVCOILLDOSDLAEVNFALPST----- 1599
DB 2393 QRSALF-----GLMEISS-----RLYT-----LLDDGLVSMQENGTSLISATLKISHKR 2436
QY 1600 -----TPPDESEVAPSPGILLATVVALNMLGELDKKKEPLTQAVGLSTQAEGRITLSLM 1656
DB 2437 KKNQPHFTTIBGIFOLFEPAANC-----DSPQVEA-----SABG--RLDTILM 2478
QY 1657 FIVNENCFYLLISQAMRYLDPVAVH--PRDKORKQELSELSTLLSS----- 1701
DB 2479 ST-----PVEIICMDVHRLRRLFLMGSSSTALSKDKKSGKPECHQ 2520
QY 1702 -----LSKFRGAGSPATGVLPSPQGSTS--LSKASPEQEPILQVQAFVRA 1750
DB 2521 DTKHTTEETVAKFLKMLASVILGKLYSEANDSQIVLSETPETLPTLLEYLKR--RN 2578
QY 1751 MQ 1752
DB 2579 LQ 2580

```

## RESULT 32

A57099 DNA-activated protein kinase, catalytic subunit - human

N/Alternate names: DNA-PK- $\alpha$

C/Species: Homo sapiens (man)

C/Date: 27-Oct-1995 #sequence, revision 27-Oct-1995 #ext\_change 09-Jul-2004

C/Accession: A57099; 159408

R/Hartley, K.O.; Geill, D.; Smith, G.C.M.; Zhang, H.; Divecha, N.; Connolly, M.A.; Admon, Cell 82, 849-856, 1995

A/Title: DNA-dependent protein kinase catalytic subunit: a relative of phosphatidylinositol

A/Reference number: A57099; MID:95401275; PMID:7671312

A/Status: translated from GB/EMBL/DBJ

A/Molecule type: DNA

A/Residues: 1789-2203 <RES>

A/Cross-references: GB:L27425; NID:9667975; PIDN:AAA79244.1; PID:9667976

C/Genetics: GDB:PRKDC

A/Cross-references: GDB:234702; OMIM:202500; OMIM:600899

A/Map position: 8q11-8q11

C/Function: involved in double-stranded DNA break repair and in V(D)J recombination

A/Description: DNA binding; DNA recombination; DNA repair; nucleus; phosphotransferase

Query Match 1.5%; Score 139.5; DB 2; Length 4096; Best local similarity 18.5%; Pred. No. 4.9; Mismatches 368; Indels 433; Gaps 56;

```

QY 354 IOLNFOYLRLLOSASGNDCTTSTACMCVGLSFLVLSLHETLGNQ----- 404
DB IFLDVF--LPRVTEALATASDQGTVAACCELLHSVMMLGKATMPBGCGGAPMYOLY 962
QY 405 ----ODIIDTACEV-----LADPSLPBLF-WGT-----EPTSGIILDSV----- 440
DB 963 KTFEVLRLACDVQVTRQLYEPLVMQILHFWFTNNKKFESQDVTSLLEALIDGVDPD 1022
QY 441 ----CGMPHLSPL-----QLRALVSGSTAKVYSF-----LDKM 475
DB 1023 STLDPCF--RCIREFLKWSIKQITPOQEKSPNNTSLPRLVSLALHPNAPKRLGAS 1079
QY 476 SFYNELVYGHKPD-----VISHED-----GTLMR-----RQTP 503
DB 1080 LAFNNIYEFREBESLVEQFPEALVYMESLALHABEKSIGTIQQCCDAIDHCRITIE 1139
QY 504 KLLVPLGGQTNLRIPQ--TVQVMDDBRAYLYVMEXYSYSWTLFTCEIEMLHVSTA 560
DB 1140 KGVSLNKKAKRRRLPRGPPPSASLCILD-----LVKM----- 1171
QY 561 DVIGHQCVKPIIDLHVIVISTDSIADCLLPITSRIMLQRLTTVISPPDVIVASCVN 620
DB 1172 -LHACGR--PQTECRHKSIELFYKFP-LIP----- 1199
QY 621 CLTVLAANPAKVMWTDLRHTGFLPVAPVSVLSQMSIAEGMN-----AGYGNLMMN 673
DB 1200 ----GNRSP-NLM-----LKQVLEBGSFLINPFGGCG-- 1230
QY 674 SEQPGEGV--VTIAELRLITLVKQGLSGTOSQGLVPCVMFVLEKMLPSYHKWRVNSGV 732
DB 1231 --QSGIAPQPTLYLR-----GPFSLQATL-C-----W----- 1256
QY 733 REQGLCLLELHAILNCHETDLSHSTPSIQFLCISLATENGQVIVINMGVDTI 792
DB 1257 ----LDLLEALR-CYNTFIFGERTVGLQVL-----GTEQSSILAAVAFLESI 1301
QY 793 DM--VMAAP-----RSDGAGGOGQGLIKTVLAASVTNNVRLKPP----- 834
DB 1302 AMHDIIAEKCFGTGAAGNRTSPQGERVYNSKTVVVRIMEFTTLTANTSBEGKLLK 1361
QY 835 ----SNVSPLEQALSQHGAGNNL--IAVLAKYIYHKHDPALPRLAIOQLKRLATVAPM 888
DB 1362 DLCTNHLRVLVQTLCEBPASIGFNIGDVQVMAH-----LPDVCNLMKAL-----KM 1408
QY 889 SVYACLGDAARDAFLTRLOSKEDEMRKIMILEFTVAVEVTPGLIEFLNLEVQD 948
DB 1409 SPY-----KQILETHLREKITAQSIIEBLCA-----VNIYGPDA 1441
QY 949 SDGSEFSLGMMSCLAHVLLELDQQDORVNCPPILHRAIAFLHLMQDRDSAMVLIR 1008
DB 1442 QVDSRLAAVSAQKQ-----LHRAGL--LHNILPSQSTDLHHSVG 1480
QY 1009 TKPKWENUTSPLFCTLSPSETSEPSILETCALIMKICILEIYVVVXGSLDQSLKDTLK 1068
DB 1481 TE-----LISLVYGIAPGBDQ-----CLP-----SIDLSGKQ-LA 1511
QY 1069 KFSIEKRAYNSGYKSL-----AHVAETEGSSCTSLLEYQMLVAMRMMLIATTH 1121
DB 1512 SGLLELAF--GGLCEBLVSLNPAVISTASIGSSGVHFSH--GEYFSLPSETIN 1568
QY 1122 ADIHLMTSVVRQFLDVLDTKALLVVPASVNCRLRGSNK-----CTLLILIR 1172
DB 1569 TELKNLDLAV--LELMOSSVDNTK--MVSALVNMLDQSFREKANOCHGQIKATITLQ 1624
QY 1173 QMKR-----ELGSVDEILGPLEILEGVLOADQOLMEKTK--AKVESAFITV----- 1217
DB 1625 HMKKDSMMAKOSPLETKAVIALAKILQIDSVSFTSHSGSPFEVFTYISLADTKL 1684
QY 1218 -LQMKEMKVSQIPQISQVLANCF-----TLOEVALALDQTRHSLALSATEDK----- 1266
DB 1685 DLHLGQAVTLLPFTSLTSGSLBELRLVLEQLIVAHFPMQSRPEPPTPRFNNVYDCKM 1744
QY 1267 --DSMETDID-----CSRSRH-----RDQDQGVCLGLHLAKELCEV 1300

```

```

DB 1745 KFLDALBELSSQPMLELMEVLCREQGHVMEELFQSSFRRLARRGSCVTQVGLLESYEM 1804
QY 1301 DEGDGSMLOVTRRLPIELFTLITL 1324
DB 1805 FRKDDPRLSFTROSGFVDRSLITLL 1828

RESULT 33
A48126
translacion activator GCN1 - yeast (Saccharomyces cerevisiae)
NAlternate names: protein G118; protein YGL195W
CSpecies: Saccharomyces cerevisiae
CDate: 21-Jan-1994 #sequence revision 18-Nov-1994 #text_change 09-Jul-2004
CAccession: A48126; S62050; S64212
R:Marion, M.J.; Crouch, D.; Himebusch, A.G.
Mol. Cell. Biol. 13, 3541-3556, 1993
A>Title: GCN1, a translational activator of GCN4 in Saccharomyces cerevisiae, is require
A:Reference number: A48126; MUID:93268304; PMID:8497269
A:Accession: A48126
A:Molecule type: DNA
A:Residues: 1-2672 <MAR>
A:Cross-references: UNIPROT:P33892; EMBL:L12467; NID:g111099; PIDN:AAA34635.1; PID:g3111.
A>Note: sequence extracted from NCBI backbone (NCBIN:132672, NCBI:P:132673)
R:Klima, R.; Coglievina, M.; Bertani, I.; Zaccaria, P.; Bruschi, C.V.
submitted to the EMBL Data Library, September 1995
A:Reference number: S62045
A:Accession: S62050
A:Molecule type: DNA
A:Residues: 1-2672 <KLI>
A:Cross-references: EMBL:X91837; NID:g1177627; PID:e203621; PID:g1177633
A:Experimental source: Strain FY1679
R:Bruschi, C.V.; Coglievina, M.; Bertani, I.; Klima, R.; Zaccaria, P.; Delneri, D.
submitted to the Protein Sequence Database, May 1996
A:Reference number: S64183
A:Accession: S64212
A:Molecule type: DNA
A:Residues: 1-2672 <BRU>
A:Cross-references: EMBL:Z72717; NID:g1322820; PID:e243797; PID:g1322821; MIPS:YGL195W
A:Experimental source: Strain S288C
C:Genetics:
A:Gene: SGD:GCN1
A:Cross-references: SGD:S0003163; MIPS:YGL195W
A:Map position: 7L
C:Keywords: translation regulation; transmembrane protein
F:69-85/Domain: transmembrane #status predicted <TM1>
F:216-232/Domain: transmembrane #status predicted <TM2>
F:929-945/Domain: transmembrane #status predicted <TM3>
F:993-1009/Domain: transmembrane #status predicted <TM4>
F:11704-1720/Domain: transmembrane #status predicted <TM5>
F:12011-2037/Domain: transmembrane #status predicted <TM6>
F:12272-2288/Domain: transmembrane #status predicted <TM7>

Query Match 1.5%; Score 138.5; DB 2; Length 2672;
Best Local Similarity 18.0%; Pred. No. 2.9;
Matches 301; Conservative 227; Mismatches 523; Indels 621; Gaps 77;

QY 326 VLAAMALLRHTLNDEFTSSVVRKIGTAIQNVFOYLTRLIQ--SLASGNDCTTSTAC- 382
DB 975 LILAMEITISVAHAEFEDPSIR-----ISIVEVLSLISLSPSKAKIARDC-FNALCQ 1025
QY 383 -----MCVGLLS-----FVLSLEHATGNODIID-----TACEVLADPS 419
DB 1026 SISVAPOEDIDMLTSLNLSLSPNQVRSST-LETLDNEPELEPMKYSPDEVFICRPDSDPS 1084
QY 420 ---LPELFWGTEPTSGIILDSVCGMFPHLSPLQLLRA-----LVSGKSTAKK 467
DB 1085 NREIADFIWEFNKRVAVNDELKSLFPLFNODSG-LRLFAANAYAFAGVSLFTSEENSSK 1143
QY 468 VYSLDKMSFYNELYKHKPHDVISHEDGTL-----WRRQ-----TPKLI--YPL 509
DB 1144 DY-LNDLINFYK--KAKPLEPILDQFGLVVASAQDQWQGRSVTAITLIKIMAKAFSA 1200

```

```

QY 510 GQGTNLRIPOGVGQVMDLDRAYLVWMEYSYSWTLFTCEIEMLHVSTADVIQHQRV 569
DB 1201 EDDIVVNIKPLVDDGGLVDEPIYQEMKRGAVLIT-----LHGONS 1245
QY 570 KPIIDLVKVI--STDLSIADCLPI-----TSRIYMLQR-LTTVISPPVDVI 615
DB 1246 KDLPFEEALSSSTDSALKENVILLYGTLARHLQQSDARHTHTIERLSTLDFPSADIQ 1305
QY 646 ASCVNCVLVLAARNAPKXWTLDRHTGFL-----PFVHAPS-----SLQOMIAGEGNA 664
DB 1306 QVNSACIPLVLFQFKQKVD--YIGILMEKLNFTVASSMKGAAMGIAGLVKGYSIA 1362
QY 665 GGYGNLNMN-----SEPOGEYVITAFRLITTL-----694
DB 1363 LSEPDILNLEAEDKKEPKRESVGFQYLSLSLCKFPPEVIELPNIKNLGDV 1422
QY 695 -----YKQUGSTQSGO--LVPCMFVLKEMLPYHKMRYNSHGR-----733
DB 1423 PEVDFATATKATKAMHTTGQVKKLIPVAVSNLDEI-----AWRTKGSVOLGNMAYL 1477
QY 734 -----EQIGLLELHAIILCHEITLHSS-----HTPSIQFLCISL-----772
DB 1478 DPTQLSASLSTIVEPIV--GVLDNSHKEVRADESLKRFGEVINPETOQLVPLVLAIG 1536
QY 773 ---AYTAG-----QT-----VINIMGIV 789
DB 1537 DPTKYTEALDSLQTOFVHYIDPSLALIHIIHRGHNSANIKRKAKCIYGNMILV 1596
QY 790 DTIMWMAAQ-----PRSGAGGQGGOLLITVYKLAASVTNNVIRLKPPS 835
DB 1597 DTKLIPYLOQLDEVEIAMVDVPNTATATARALGALVEHLGEOF-----P 1644
QY 836 NVNSPLEQALSOHGANGNL--IAVLAKYI--YHKIDPALPLAIQLKRLATVAPMS 889
DB 1645 DLIRLDLTLSDESKSGRLSSAQALAEVIGLITKLDENLPT-----LGVNFR 1697
QY 890 VY-----ACLGNDAA--AIRDAFLRLQSKIEDRI-----KXMIIEFL 926
DB 1698 AYIREGFWPLLLFLPVCFSQGFAPYINQIQLPILSLANDENIRDALKAGKLIYKVA 1757
QY 927 TVAV-----ETQPLIE-----LF-----LMEVKGSDSGSKERSLG 958
DB 1758 TKAVDLLPELERGMFENDRIKLSVOLTEGELLFOVTGSSRNEFSEEDDHGERS-- 1815
QY 959 MMSCLHNVLELIDSOQODRWYCPPLHRAAIAFLHALMODRDSAMVLRTKPFEMENLT 1018
DB 1816 -----GKLVDYLGQDRDR-----LALFLVCNDSIGVIRATVDYWK-- 1854
QY 1019 SPLFGTISPPSETEPSTLETICALIMKICLETIYVVKGSIDQSLKDTLKKFSTIEKRPAY 1078
DB 1855 -----ALVPNTPRAVKEILPTLQGMIVTHLASSNVAFINIAOQLGDLVRRVGNA-- 1905
QY 1079 MSGVYKSLAVHVAETEGSS-----CTSLLEYQMVSVA-----WRMLLIATTHADI 1124
DB 1906 LSQILPILSESLIETNSDSRQGVCIAL--YELIESASTETISQPOSTIVILITLALID- 1962
QY 1125 MHLTDSVVR-----QLFLDVLDTKALILVPAVNCRLSGMKCTLLILLRQMKREL 1178
DB 1963 ---ESATVREAAALSPFQDVVGKT-----1985
QY 1179 GSVDEILIGPLTEILEGVLOAD-----QOMETKAKVPSAFITVL-----1218
DB 1986 -ADEVLPYLHMLHSSDNDSPALLGQEIWSKSDVIFPLILFTLLAPPIDAFRASALG 2044
QY 1219 QMKEMKVSDFPOYSQVLVNVCELTQOEVIALFDOTRHSIALGATEDKDS-----METDD 1273
DB 2045 SLAEVASSALYKRSIIIN-----ALVD-----AIGTSBESSTGALIELLDR 2088
QY 1274 CSRRHARDQDGVCVLGLH-LAKELCEVEDBDGSMLOVTRRLPTLPTLLTLLLEVSLMKQ 1332
DB 2089 VFLSVNDE-----GLHPLLOQIMSLKSDN-----IEKRIATLEKLPNPFKVTV-LDF 2136
QY 1333 NHHTEATLHLLLTARTQOGATVAVAGAGITQISICLPLLSVYQULSTNGTQTPSARKSL 1392

```

```

DB 2137 DVYIPNVSAIISLDEDDR-----VNGNNALSTLLKKV 2173
QY 1393 DABWPGVYRLSWSLMEQLLKTLRYNFLPEALDFGVGHORTLOCLNAVTVOSLACEE 1452
DB 2174 DKPT-----LEKIVKPAK-----QSLALYGR 2194
QY 1453 ADHTVGPILQSNFMKEMHFLPOLMRDIOVNIQVLCQACTSILHSKMKLQHYLQNGND 1512
DB 2195 QGQDVA-----AKLPR-----GPNCVL-----PIFLHGLMGNSD 2225
QY 1513 GLPSAVAQVQRPSSAASAPSSSKOPADTE-----ASEQALHTV 1554
DB 2226 -----ERESALAIADVSKTPPANLKEFVYITGPILRVNGERRSSDIKA--AI 2273
QY 1555 QYGLKLSTKTALALRHFTPDVCQIILDSGLDAEYNFLFALS-FTTPFPDSVADSPGT 1613
DB 2274 LFLNLVLFIKPMFLPFLPOL-----QRTFVKSLSDATNETLRLRAAKALGA 2321
QY 1614 LLAFTVVALMGLDLKKKEPLT-QAVGLSTQA--EGTRT--LKSILMTMB 1660
DB 2322 LI-----EHQRPVDELVELVTGAKQATDGVKTAMKALLEVIWK 2362

RESULT 34
S46715
hypothetical protein YHR099w - yeast (Saccharomyces cerevisiae)
N/A:alternate names: hypothetical protein H9332.1
C/Species: Saccharomyces cerevisiae
C/Date: 28-Oct-1994 #sequence_rev150n 28-Oct-1994 #text_change 09-Jul-2004
C/Accession: S46715
R/Vaudin, M.
submitted to the EMBL Data Library, May 1994
A/Description: The sequence of S. cerevisiae cosmid 9332.
A/Reference number: S46715
A/Accession: S46715
A/Molecule type: DNA
A/Residues: 1,3744 <VAV>
A/Cross-references: UNIPROT:P38811; EMBL:U00060; NID:g487928; PIDN:AB68923.1; PID:g4879;
C/Genetic:
A/Gene: SGD:TRAI
A/Cross-references: SGD:S0001141; MIPS:YHR099w
A/Map position: 8R

Query Match 1.5%; Score 138.5; DB 2; Length 3744;
Best Local Similarity 17.6%; Pred No. 5;
Matches 252; Conservative 214; Mismatches 492; Indels 477; Gaps 61;

QY 100 YLQEDYGRTRDSYKTVLOD-----EROSQALILKIADYYEERTC 139
DB 317 FLQDYNVFPBULIRLQDCPSLSARKKELLHATRHILSTNYKKLFLPKLDYLPDERIL 376
QY 140 I-----LRCYLHLITTFQDERHPYRVEYADCYDKLEKEIVSKYRQOPPELYKTEAFTWET 194
DB 377 IGGFTMHETLRLAY-----STVADFHNIRSEL-----QLSEIEKTX 414
QY 195 HGNLMTERQVSRFVQCLREOSMLLEIFLYAAVFEMASDILLVKMKKEQFGSRQTN 254
DB 415 -----IKIYGYLDESLALTVOJMSAKLLNVERILIKKENPQDEAPRAKKLL 464
QY 255 RHLVDETMDFVDRIGYFSAIILVSGM-----DISSLHAKALDRRELHOPADGLIC 307
DB 465 MLIIDSYNRKFTLNQYDTIMKYGRVETHKKERKAKNSIQDNKDESEFMRKVLPR 524
QY 308 QMDDCM-----LTFGDIPIHAPVTLAALRLHTLNPEE 341
DB 525 SDDHLMFPQPKEDINDSPDVMTESDKVKNQDVEMFDIKYAPILL-----PTP 575
QY 342 TSSVAKIGGTALQNLVFOYLTRLOSLAS-----GANDCTTSACM-----383
DB 576 TNDPIK-----DAF-YLYRTLMSPLKTIHDLKVNPPNPNETVANNPKLMAVSXRV 625
QY 384 -----CVYGLLSFV--LTSLELHTLGNODIIDTACEVLADSLPELFWG 426

```

```

Db      626 FSYBEVIVKDLPHRECIIGLKPFKDHNEKSPETTKGHFDI-----SNRSLP----- 672
Qy      427 TEPISGIGIIDSVCNMFPHLLSPLOLRALVSGSKTAKKVVYFELDKM-----SFINE 480
Db      673 -----VSATKDARELMYLAEMFMQMDNAETNE 700
Qy      481 LYKHK---PHDVISHEDGTLMRRQ---TPKLLYVGGGTNRIRIQG---TVGOVWMDRA 531
Db      701 IIEBELPEVYERMEDESLHVAQSFLTSETSPNFAGILRFKGLKDKLGNVDFTSN 760
Qy      532 YLVN-WEXYSYSSWTLF--TCEIEMILHVSTADVIQHCORVAPRIIDLVHKVISTDLSD 588
Db      761 VLILFLKLSFMSVNLFPVINEVLLPH-----LNDILINSKXSTTAAE 804
Qy      589 CLLPITSRIYML-----ORLTVISPPVDVIVASCNCLTVLAARP--AKV 633
Db      805 PL-----VVFYLRITLFRSIGGFENLYRSIKRILQVLLQSLNQW--ILTARLPHEREL 857
Qy      634 WTDLRHT-----GFLPVAHP-VSSLSQM--ISAEQMA----- 664
Db      858 YVELCTIVPRLSVIAPLPFLMKPLVPAIQYVDPVLSQGLRTIELCIDNLTAEYFDPIT 917
Qy      665 -----GCVGNLNMNSEQPOG-EYGVIAFLRITTLVKGQLGSTOSQGLVPCVMEVLKEML 719
Db      918 EPVIDVSKALFNLLQPPFNHAIASHNVRL-----QKLGRRKQFLKRPDTDLTEKTEL 972
Qy      720 PSYHKRWYNSHGVREQIGCLILELILHALNL--CHETDLSHSHPSLOFLCISLA--- 773
Db      973 DIDAIADFKINGMEDVPLSVTPGQSALNIIQSYKSDIHVKSAYKVLTCVLLMTKXS 1032
Qy      774 -----YTBAGQYINIMGIGVDITDMMAQPSDGEQGGQGLIKTYKLAFSVY-- 825
Db      1033 AEPFTNTTELTKTIVNSIKERIGIEKNFDEPYNKDYSGNQLFLRLLESVFYASTI 1092
Qy      826 -----NNV-----IRLKP-----SNVVSPLERQA 844
Db      1093 KELDDAMDNLNLDHFCLLQWNTTLNKNYNTFIDIKNPMFLDSSLIIDAIAPFA 1152
Qy      845 LSGHAGNNLILAVLAKTYHK-----HDPALPRLAIQL-----KR-- 881
Db      1153 LSYIIPREVGVLAKKRIYKESCLIGEELASHSFIPELAKQIHLCYDETYVNRKG 1212
Qy      882 -----LATVAPMSVVA-----CLGNDAAIRDAFLRLQSKIEDMKIMILEFLVA 929
Db      1213 VLGIRVLIDNVKSSSVFLKTYQYNLANGLFLVKDTQSEAPSAITDSAEKLLI--DLISIT 1271
Qy      930 VETQPGILIEFLINLEVKGDSGSKESFSLGMSCLHVALELIDSOQODRYWCPPLLHRAI 989
Db      1272 F-----ADVKEEDLGNKYLEN---TLTDIVGELSNANPKYVNAQOKSLH--TI 1314
Qy      990 AFLHALMODRRDSAMLVLRTRKPKWMENTSPFLGTLSPSETSEPSILETCALIMKI-- 1046
Db      1315 SNL-----TGPIYKLMDSHKOPLSPIF-----AKP--LRALPFTMOIGNV 1354
Qy      1047 -----ICL-----EYIVVKGSL-----DQSLKOTLKKS-----IERFA 1077
Db      1355 DAITFCISLPNTFLTFLNEBELFRLLQESIVLADABESISTNIOKTEXSTSEQLVQLIA 1414
Qy      1078 YMSGVKSIAVVAETE-GSSCTSLLEYOMLVASARMMLIATTHADIMHLLTDSVVRQL 1136
Db      1415 -----CIKLALAIKNEEPATAOQGNIRIRILAVFKTML--KTSPEIINTTYALKKSL 1467
Qy      1137 FLD-----VLDTGKALLLVPAVNCRLGSMKC--TLILLILNQMKRELSVDEILGP 1187
Db      1468 AENSLKPEKRLONGKPLMLNLSHDQKLTVPGLDLSKLELLIAYFKVEIGR--KLDDH 1525
Qy      1188 LT-----EILEGVLADQQLMEKTKAPPSAFITVLQMKEMKVSIIPOYSQVLVN 1237
Db      1526 LTAMCRVAVLDLTLF--GODLAEQMPTKIIVSIINIIFH-----LPQADMFVN 1571

```

```

509811
hypochemical protein U48 - human cytomegalovirus (strain AD169)
N:Alternate names: hypochemical protein HFRFO
C:Species: human cytomegalovirus, human herpesvirus 5
A:Note: host Homo sapiens (man)
C:Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 09-Jul-2004
C:Accession: 509811
R:Chee, M.S.; Bankier, A.T.; Beck, S.; Bohm, R.; Brown, C.M.; Cerny, R.; Horenell, T.;
M., Burrell, B.G.
Curr. Top. Microbiol. Immunol. 154, 125-169, 1990
A:Title: Analysis of the protein-coding content of the sequence of human cytomegalovirus
A:Reference number: 509749; MUID:90269039; PMID:2161319
A:Accession: 509811
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-2241 (CHE-)
A:Cross-references: UNIPROT:P16785; EMBL:X17403; NID:G59591; PIDN:CAA35407.1; PID:gl7808;
A:Note: this sequence was submitted to the EMBL Data Library, December 1989

Query Match 1.5% Score 138; DB 2; Length 2241;
Best Local Similarity 19.6%; Pred. No. 2.3;
Matches 295; Conservative 190; Mismatches 529; Indels 494; Gaps 71;

Qy      398 LHTLGNQODIIDTACEVLADEPSLEL-----FW-----GTE-----PTSGIGIIL 437
Db      341 LPALGRYQQLVDEVEGELKALTLPPLPANTSAMTLHAAGTESGANAATATAPSFDEAFLT 400
Qy      438 DSVCGMFPHLI-----SPLLQILRALVSGSKTAKKYYSFELDKKSFNYELKX 484
Db      401 DRLQQLIIHVNQRCCLRRPCGPOSAQAQAVRAYL---GLSKLDAFLNWN-LHHGLDLQ 456
Qy      485 KPHDVISHEDGTLMRRQTPKLLYPLGGGTNRIRIQGTGVQWMDRAVLWEXYSYSW- 543
Db      457 RMHDYLSHKT-----TKGYSTIDRALLEMQVVPDPYG 490
Qy      544 -----TLFTCEIEMILHVVS--TADVIQHCORVAPRIIDLVHKVISTDLSD---CLLP 593
Db      491 RQHGPALIAWEENLRVYESKPTWELSGRLQRF-----YKRPMPVDSVFCRLP 541
Qy      594 TSRITMLQRLTITYISP-----VDIVASCNCLTVLA 626
Db      542 D-----FORLTQVIEGRVRLQORREYHGVYHLAGLITSIDIHLDADSLRRRELTK 595
Qy      627 ARNPA-----KYMTDL-----RH--TGFLPFAHPVPSISQ 655
Db      596 ALQPLDNDNAKQELRRLGNAKMLBELQMDLDRISTQULTRVANHILANGFL-----PEVDLKQ 650
Qy      656 MISAEQWAGGYGNILNMNSEQPOGEYGVIAFLRLITTLVKGQL-GSTOSQGLVPCVMEV 714
Db      651 M-----ENVVEQVLFLFYDLRDLKLCDSGYEGFV-----V 681
Qy      715 LKEMLPSTYHKRWYNSHGVREQIGCL--ILELILHALNLCHETDLSHSHPSLOFLCISL 772
Db      682 IREQL-SY-----LMTGVTRDNPVPLLOEILQJRHAYVQOATQONERLTOIHDLHVI 732
Qy      773 AYTBAGQYINIMGIGVDITMWA-----APRSDDGAEQGGQGLIKTYVLAFSVTNNV 828
Db      733 -----ETLVADPESGSALTALVQEQALDELAGGLQLEVDQORLQNALAS----- 781
Qy      829 IRLKPPSNVVS-----PLEQALSGHAGNNLILAVLAKTYIYHKHDPALPRLAIQ 877
Db      782 -RLVBEHEETORFLDGLSYDDPPNEQITKR-----POLREM 817
Qy      878 LLK-----RLATV-APMSVYACLGNDAA--AIRDAFLRLQSKIEDMKIKYMIILEFLVA 929
Db      818 LRRDEQFTLRILINAVLSWFTLVNRLARDESPRTFFDAVSLILQOLPPDSHEREDLPA 877
Qy      930 VETQPGILIEFLINLEVKGDSGSKESFSLGMSCLHVALELIDSOQODRYWCPPLLHRA- 988
Db      878 NATYAQWVKLEQIEKAGTGASEKF-----QALRELVYFRNNEYFFQHWGRIGV 929
Qy      989 ---IAFLHALMODRRDSAMLVLRTRKPKWMENTSPFLGTLSPSETSEPSILETCALIMK 1045

```

Db 930 GPVTELYERYOHENEEOH--LERLERMOBAGRLVLT-----SVEDQVRLAR 977  
 Qy 1046 IICLEIYVYVKGSLDQSLKDTL---KKPSIEKRFAYVSGVYSU---AVHVAETEGSSCT 1099  
 Db 978 APSHRVHMQOQTLTKKQMDPLDKKQKQOEQROLDGYQKQOQDLQRFVADVAKGEML 1037  
 Qy 1100 SLELYOMLVSAAMRL-----IATTHADIMLTDSVVRQLFLDVLDTGATALLVP 1151  
 Db 1038 STIPHQPLEATLLELLGLDQRAQPLDKFNQDLISALQOLSK-----LDG----- 1083  
 Qy 1152 ASVN-CLR---LGSN--KCTLLILLRQWKELSGVDEILQPLEILE-----GVLOA 1198  
 Db 1084 -RINECHLGVLTGVRRCHPRERAMQTOASLNLHLDQILGPOLLINHTOQALQHAHVAQA 1142  
 Qy 1199 DQCMETKAKVFAFIVLQMKEMKVSIDIPOYSQVLVNCETLOEVIATLFDQTRSLA 1258  
 Db 1143 --QTEKQOQDPTTATIGSEFE---GDPARYSSQCKMEEQLOE-----TROQMT 1188  
 Qy 1259 LGSATEDKDSMETDSCSRHRDQDGVCLGLHLAKELCEVDBDQSGMLQVTRRLPLP 1318  
 Db 1189 ETSERLDR-SLRQDPGSSSVTRVPEKP--FKGQELAGRITPPRADFQO-----PVFK 1237  
 Qy 1319 TL-----LTTLEVSILMKQNLHTEATLHLLTLTARTQCATRVAAG 1361  
 Db 1238 TLLDQADARKALSDADLLNQKVQTLRQ-----RDEQLSTA----- 1276  
 Qy 1362 ITQSLICPLSVOLSTNGTQTPSASRKSIDAPSPGVRLMSLMELKTLRYNPLP 1421  
 Db 1277 --QVLMTDLVTRHMGSGLDVTTTPDAK-----ALMEKPLETIR----- 1312  
 Qy 1422 EALDPGVGHQERTLOCLNAVTVQ-SLACLEADHTVGFIQLS-----NFMKEWP-H 1473  
 Db 1313 ---ELIGKATQO-LPYLSAERTVWMLAFLEEA-----LAQITADPTPHHSGRTYRN 1362  
 Qy 1474 LPOLMRDIOVNLGYLQO---ACTSILHSRKNLQNYLQNKNGDGR-----SAVAGRV 1522  
 Db 1363 LQOQAVESAATVLAQIQEONNAACENPT---AQHOEATANASTPRYDMQAVEAWQRL 1417  
 Qy 1523 QRPSSAASAPSSSKOPADTEASEBOQLHTVQYGLKILSKTALAALHFTPDVQCQLLD 1582  
 Db 1418 E-PERVAVGA-----ARHQKQVELLQRLQGLDLE-----LQ 1449  
 Qy 1583 QSLDLAEY-----NFLPLASTTP-----TPDSEVAPSPGILLATVNV----- 1620  
 Db 1450 ETL-ATEYFALLHGIQTFYSGLDFRSOLEKIRDLRTFPAELAKRRGRLSNEGVLPPNRK 1508  
 Qy 1621 --ALNMLGELDKKKEPRLQAVGLSTQAEGRITLKSILMFTMENCFYLLISGAMRYLADPA 1678  
 Db 1509 PQAATTSIGAFTRGLNALERHVLQGHQYILANKLNGSSLVYRLDIPSVL---PATHTDP 1565  
 Qy 1679 VHPRDQKOR 1686  
 Db 1566 LHMDRRLR 1573  
 RESULT 36  
 T13949  
 neurofibrin - fruit fly (Drosophila melanogaster)  
 C:Species: Drosophila melanogaster  
 C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 09-Jul-2004  
 C:Accession: T13949  
 R:Title, I.: Hamidgan, G.E.; Cowley, G.S.; Reginald, S.; Zhong, Y.; Gusejla, J.F.; Harihat  
 Science 276, 791-794, 1997  
 A:Title: Rescue of a Drosophila Nf1 mutant phenotype by protein kinase A.  
 A:Reference number: Z17826; MUID:97277221; PMID:9115203  
 A:Accession: T13949  
 A:Status: Preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-2764 <THE>  
 A:Cross-references: UNIPROT:O01399; EMBL:L26502; NID:g1929432; PID:g1929433; PIDN:AAB585  
 C:Genetics:  
 A:Gene: Nf1  
 A:Cross-references: FlyBase:FBgn0015269

Query Match 1.5%; Score 137; DB 2; Length 2764;  
 Best Local Similarity 18.8%; Pred. No. 3.9; Mismatches 627; Indels 602; Gaps 93;  
 Matches 340; Conservative 243;

Qy 169 LEKELVSKYQOEELKYTEAPTW-ETHGNIMTERQVSRMFVQCLRQSMLEIIPUYA 227  
 Db 755 LQKRIITLRL-KLEHCVHGVPAWMEETFRMWEVSSKYLQYRPKCKGSDQ----- 803  
 Qy 228 YFEMAPBDLVITKMFQEGFSR---QTRNLVDETMPPVDRIQYFPAALLIVEGMDI 283  
 Db 804 -----AEVF-HRWGKRRASSHQSEHDLQEIEMAMMTFLALGCV----- 845  
 Qy 284 ESILKCALDRRELHOPADGLICQMDLMTFFGDI-PHAPVLLAWALLRHTLNPEETS 343  
 Db 846 -CHKRS-SSRQMLQOSQNN-----ASLG-----SLQNSLYSSSTS 881  
 Qy 344 SVAKIGGTALQLN-----VFQYLRLLQSLASG-----G 373  
 Db 882 SGHSLHPSIVSLSTLPPAPQDVSYCPVQFVGLRLRLVCSNEKILGNIQKVKELVG 941  
 Qy 374 NDCITSTACMCYGL-----SVLVSLEHTLGNOODIIDTA 411  
 Db 942 EEMSTQ---LVPILFDQVRAIVEKFFDQGGVNVNVTIDINTQFIHTIYIMKSIDPK 996  
 Qy 412 CEVLADPSLPELFWGTEPTSGLGIIIDSVCGMF-----PHL-----LSPILQL 455  
 Db 997 AN--KQPNNDQ-----PSPSEHLGV--TSIEGMVLGIVRYVRHIDMTYAIRIKTKQLV 1048  
 Qy 456 RALVSGKSTAKKQVYSLDKKSPFYNELYKHPHDVISHEDGTLNRQOTPKLLYPLGGO-- 512  
 Db 1049 EVMMKRKRD---LAFQOEKKFRNKLVEYL-----TDWVGTSQHIAPSSADAA 1094  
 Qy 513 --TNLRIPQGTGVQWMLDRAVLV-----RMEVYSWMLFTFCIEMLHNVST 559  
 Db 1095 ILNTSILIPFDLQACHEAVALLRGLPLOPESDRKDLMDASALFKFTYLPMLNLN- 1153  
 Qy 560 ADVIYHCQVKKPIIDLVHKVISTDLSTADCLPITSRYMLQRLTVVISPPVDVIASCV 619  
 Db 1154 -DCIDSEBAKEMNNP-----LPPRRP--WAACKLTALRN----- 1187  
 Qy 620 NCITVLAAKN-PAKYWTDLRHT--GFLEPV--AHVSSLSQMS-----AEGNN 663  
 Db 1188 --ATILAMSNLIGANIDSGIMSIDIGYNPDLOTRAAFMEVLQIILQOQTEPFTLAEVL 1245  
 Qy 664 AGGQGNL--LMNSEQOGGEVGTIAFLRITTLTVKQGLSTQSGVPCWPFVKEML-P 720  
 Db 1246 ADRFEQVLQVLTVMISDKGELPIMALLANVVTISQMDLRL-----VLVTLFPAKHLISP 1299  
 Qy 721 SYHKWRYNSHGVREOIGCLILBELIHALILNCHETDLHSSHTPSLOFLCISLAYTEAGQT 780  
 Db 1300 LMMNPFREVEVSD-----CMQT-LFRGNSLSGKIMAFCKIY----- 1336  
 Qy 781 VINIMGIGVDTIMVMAAPRSDGAGQGGQGLLITVKLAESVTNNVITLKPSSNVSP 840  
 Db 1337 -----GASYLQMLLEPLLR-----PLDDEBEETCFEV--DPAFLDPEEDT--- 1374  
 Qy 841 LEOALSGHAGNNLAVLAKY---LYHKHDPALPRLAIG-----LYRLATVAPMSV 890  
 Db 1375 -EQ-----HRNNLIALTKVVDALINSSDLRPLQRLRSCHCLYVLSKRFPNL----- 1421  
 Qy 891 YACIGNDAAIRDAFLTRL-----OSKIDMEIKWMLLEFLTVAVE 931  
 Db 1422 --LQNNIGAVGVIFLRFINPAIVEPQELGIVDKOVHSAKKGMLMSKILQNIANHVE 1478  
 Qy 932 TQGGLELFLNLEVKQSGDSKESFSGMS--CHAVLELIDSOQDQRYMCPRLHRAATA 990  
 Db 1479 PSKEQMLCFNDPLRDHFEGRRFFQIASDC-----ETVDQTSQSN---SPISDANVL 1529  
 Qy 991 FLH-AW-----QDRDQAMVLVLRKPKFEMENLTPRLFGTSLSPSETSEPTILET 1039  
 Db 1530 ALHRLIMTHQEKIGDVLSSRDHKAVGRRPFDKM-----ATLALYLGPRPHKRVDSHMF 1584

```

QY 1040 CALIMKICILEYVVKSGLDQSLKD-----TLKFSI-----EKFA 1077
Db 1585 SS-----YARMSIDMSSTNFEIINVKQMKHEKPKILKSNINIIYQAGTSSG 1633
QY 1078 YMSGVKSIAVNAVEEGSSCTSLIEYQMLV-----SARMLLIATTHADIMHILTDS 1130
Db 1634 YPVYVYIARRYKIGETNG-----DLIIYHVIITLKPCFCSPEEVIDPHTCSIDNRFRTPE 1689
QY 1131 VVRQQLPDLV--DOTKALLLVPAVNCRLRGSMKCTLLILLNQMKELGSV--DELGP 1187
Db 1690 LQKMFYVLPVAVENVAHVI--YNC-----NSWREYTKPHRILAP 1730
QY 1188 L-----TEILEGVLOADOQ-----LMEKTKAKVFSAFI----- 1215
Db 1731 LKGRKLLFLESFPMKLTDFIDAEQKLPGLSLDEDLKFSNALKSHKDTKALIKGP 1790
QY 1216 TVLQMKEMKAVSDIPOYSQVLINVCETLOEVIYALFDQRHSLALGATEDKDSMETDGS 1275
Db 1791 TALQITSAEKTAVLAHSLVTLNDVYVASEIEEVCVLDVNDQFTLSITNES--GQLSFIHND- 1848
QY 1276 RSRHRDQDGVCGVGLHAK--ELCEVEDGDSMLQVTRRLPILP-----TLTTLBVSIL 1328
Db 1849 -----DNIVQAIHIRNWEISQPD-----SVTVHOKIRPKDVPGLTLNMLNLT 1893
QY 1329 -RMKONLHFEATLHLLTLARTOQATAVAGAGI--TOSICLPILSYVQLSTNGTAQTPS 1386
Db 1894 GSCQPNLR--TNAVNLCAIATATPD--LKIEGQLLETGGLCP-----SNNTIIFKS 1941
QY 1387 ASRKSLDAPSWPGVYRISMSIMEQLLKTLYNPLPEALDFVGHQERTLQCLNAVTVQS 1446
Db 1942 VSEKL--ATNEP--HLTFLEESIQGF-----QRTTIELKH----- 1974
QY 1447 LACIEADHVTGFIQLQSNFKEMWFHLPLMRDIQVNLGYLCAQTSLLHSRKLQHYL 1506
Db 1975 -LCLE-----YMTPW-----LKNLVKCKSNDKSKLKYSQI 2005
QY 1507 QNKKGDCIPSAVQVRQPPAASAPASSKOPADTASEQALH--TVQYGL-----LKI 1561
Db 2006 LDK-----LINLTIDQKEMYPVQAKIWGSIQIPELIDMWLDNPLHKSITTYGGLSPQEI 2061
QY 1562 LSKTLAAL-----RHFTPDVQQL-----LDQSL--DLAEY--NFLFALSF 1598
Db 2062 MADTAVALASANVOLVSKVITRICRVMDKSCCTNPQYLBQHMMDIDAIIGRYLMLSF 2121
QY 1599 TTPFPDSEVAPSGTILATV-----INVALMNLGSLDKKKEPLTOAVLS 1642
Db 2122 NNCL--DVATSVPYLFTHTITFLVCSGSPSMKASTHGLVINIHSILCTNP-----S 2171
QY 1643 TQAGSTRILK--SLMFTMENCEFYLLI-----SOMARYLDPVAVPRDK-----ORMKQL 1691
Db 2172 FSEERQVRLSLDEFSLPK--FYLLFGISKVKSAAVTAFRSSCRHPTDKMLGNEKVTOP 2230
QY 1692 SSELSTL--LSGL 1702
Db 2231 PADRERLSLPSL 2242

```

```

RESULT 37
T13947
neurofibromin - fruit fly (Drosophila melanogaster)
C/Species: Drosophila melanogaster
C/Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jul-2004
C/Accession: T13947
R/Title: I.; Hamilgan, G.E.; Cowley, G.S.; Reginald, S.; Zhong, Y.; Gueella, J.F.; Harihar
Science 276, 791-794, 1997
A/Title: Rescue of a Drosophila Nf1 mutant phenotype by protein kinase A.
A/Reference number: Z17826; MUID:97277221; PMID:915203
A/Accession: T13947
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Reids: 1-2802 <THE>
A/Cross-references: UNIPROT:O01398; EMBL:L26501; NID:g1929430; PID:g1929431; PIDN:AB589
C/Genetics:

```

A:Gene: NF1  
A:Cross-references: FlyBase:FBgn0015269

Query Match 1.5%; Score 137; DB 2; Length 2802;  
Best Local Similarity 18.8%; Pred. No. 4;  
Matches 340; Conservative 243; Mismatches 627; Indels 602; Gaps 93;

```

QY 169 LEKELSVKROQFELYKTEAPTW--ETHGNMTERQVSRWFVQCLREOSMLLEIFLYYA 227
Db 755 LQKIMTLIR--KIECHAGVUPAMEETFRWESSKYLQYTPKKGEQGQ----- 803
QY 228 YFENAPSDLLVLTQMFKEQFGSR-----QTRHLVDETMDFVDRIGYFALLIVEGMDI 283
Db 804 -----AAVF--HRGKGRKRAHQSEHLEQINEMAMWTFLLAGV----- 845
QY 284 ESLKCALDRRELHQRQDGLICQDMDCMLTGGDIPHAAPVLLANLHNTLNPERTS 343
Db 846 -CLHRS--SRQMLLOOSQNN-----ASLG-----SLAQNSLYSSSTS 881
QY 344 SVARKIGGTALQLN-----VFOYLRLLQSLASG-----G 373
Db 882 SGHSLHPSTVSLSTLPAPAPQDVSCPVTQFVQQLRLVCSNEKIGLNQKVKELVG 941
QY 374 NDCYTAQCMCVGL-----SFVLTSLLEHTLGNQDIIIDTA 411
Db 942 EEMSTQ-----LYPILFDQVRAIVEKFPDQGGVNNVNTDINTQFIHTIYIMKSLDPK 996
QY 412 CEVLADSLPELFMGTEPSTGLIILDSVCMF-----PHL-----LSPLQLL 455
Db 997 AN--KDPNNDQ-----PSPSEHLAGV--TSIEGMLGIYRVYHLLDMTYAARIKTKLCQLV 1048
QY 456 RALVSGSTAKKVSFLDKMSFYNELYKHKHVDVISHEDGTLMRPQPKLYPLGGQ--- 512
Db 1049 EVMKMRKD--LAFRQEMFRKKLVEYL-----TDVWNGISHQIAPSSADAA 1094
QY 513 --TNRIPOGVGVMLDDRVLV-----RMEYSYSWTLFTCEIEMLHVVST 559
Db 1095 ILTNTSLIFRDLQACAEVALLRGLPLQPEESDROGLMDAKSALFLKYFTLFWNLIN- 1153
QY 560 ADVIQCORVPEIIDLHKVISTDLSTADCLPITSRLYMLQLLTYVISPVDVIASCV 619
Db 1154 -DCIDSSAEKEMNTP-----LLPRPR--MAAGKLTALRN----- 1187
QY 620 NCLTVLAARN--PAKWTMDLHNT--GFLPV--AHVSSLSQMIS-----AEGMN 663
Db 1188 --ATILANSNLDGANIDSGLHSHSIDLGNPDLQTRAFMEVLTLLQOGTEFDTLAEIVL 1245
QY 664 AGGYGNL--LMNSEQPOGEYGVITAFRLITTLVKGQIGSTOSQGLVPCVMFVLKEML--P 720
Db 1246 ADRFQVLQVLTMTSDKGLPIAMALANVVTTSCMDLAR-----VLVTLFDKHLILSP 1299
QY 721 SYHKRRYNSHGVRQICGLIELIHAIILNCHETDLHSHSTPSLOFLCTSLAYTEAQT 780
Db 1300 LLMNMFYREVEVSD-----CMQF--LFRGNSLGSKIMAFCKIY----- 1336
QY 781 VINIMGIGVDIIDWMAAQPRSDGABEGQGGQLIKTYKLAFSVTNNVIRLKPSNVVSP 840
Db 1337 -----GASYLQMLLEPLIR-----PLDDEBEETCFEV--DEARLDPTEDL--- 1374
QY 841 LEQALSGHAGNNILVATKY--IYHKDPALPRLAIQ-----LKLKATVAPMSV 890
Db 1375 -EQ-----HRNNLIALTQKVPAIINSSRLPQLRSMCHCLUYLSKRFPLN----- 1421
QY 891 YACGNDAAIRDAFLTRL-----QSKIEDMRKIKVMLLEFLTAVE 931
Db 1422 --LQNNIGAAGTVIFLRFINPAIVSPQELGVDKOVHSSKRGIMLSKILQNIANHE 1478
QY 932 TQPLGIELFNLKVDSDSGSKESLGMWS--CLHAYELIDISQGDQDRWCPLLHRAIA 990
Db 1479 FSKQGHMCFNDPLRDFEAGRRFFTOASCC-----ETVDQTSMS--SFLSDAVL 1529
QY 991 FLH--ALW-----QDRDSAMLVLRTKPKWEMNTLSPFLGTLSPSETSPSILLET 1039

```

Db 1530 ALHRLMTHQKIGDYSSSRDHKAVGRPFDM-----ATLLAVLGPEHKKPVDSHMF 1584  
 Qy 1040 CALMKIKICIEIYVVGSLDQSLKD-----TLKKRSI-----EKRA 1077  
 Db 1585 SS-----YARSSIMSSSTNPEEINVKQMEKEEFTLKSNNITYQAGTSKSG 1633  
 Qy 1078 YMSGVSLAVVAETEGSSCTSLLEYQMLV-----SARMLLIATTHADIMHLLDS 1130  
 Db 1634 YPVYIYARRYKIGETNG-----DLIYHVIITLKPFCHSPPEVIDFHTGSDNRFRREF 1689  
 Qy 1131 VVRQQLFDVL--DGTALLVLPASVNCRLGSKKCTILLILLOMKRELGSV--DEILGP 1187  
 Db 1690 LQKMFYVLPVAYENVAHYI-----YNC-----NSWREYTKFPHRIIAP 1730  
 Qy 1188 L-----TEILGVLQADQO-----LMEKTKAVPSAFI----- 1215  
 Db 1731 LKGRKLLFLESFNKLTDFIDBQOKLFGATLSLDEDLKVSNAIKSHKOTKVAIKGP 1790  
 Qy 1216 TVLOMKEMKASDIPOYSQVLNVCELTLOEVIYALFDQTRHSIALGSATEDKDSMETDGS 1275  
 Db 1791 TALQITSAEKTQVLAHSLVLDVYVYASIEVCLVDNQFLTSTNES--GQLSFIHND- 1848  
 Qy 1276 RSRHRDQDVCVGLHLAK--ELCEVDEGDSMLQVTRLPILP-----TLTTLLEYSL 1328  
 Db 1849 -----DNIVQAIHIRNEMELSQPD-----SVVHOKIRPKQVPGTILMMALLNL 1893  
 Qy 1329 -RMKONLHFEATLHLTLTARTQOGATAVAGAGI--TQSICLPLSYVQSTNGTAQPS 1386  
 Db 1894 GSCPNLR--TAANRLCALATPD--LKIGQLLETQGLCP-----SNNTIFIKS 1941  
 Qy 1387 ASRSLDAPSWPGYVRLSMSIMEQLKTLRYNPLPEALDFVGHQERTLOCLNAVRTQS 1446  
 Db 1942 VSEKL--ATNEP--HLTFLEBSIQGF-----QRTTIELKH----- 1974  
 Qy 1447 LACIEADHTVGFILQSNFMKEWHFHLPOLMRDIQVNLGILCAQCSLHRSKMLQHYL 1506  
 Db 1975 -LCLF-----YMTFW-----LKNLVFCKSNDKSKLKXVQI 2005  
 Qy 1507 ONKKGDLPSVAORVQRPASASAPSSKOPADTEASEQALH--TVQYGL-----LKI 1561  
 Db 2006 LDK-----LINTIQKEMYPVQAKIMSGIQIPELIDMVDNLFHLSITYGGLSPQVEI 2061  
 Qy 1562 LSKTLAL-----RHFTPDVCOIL-----LDQSL-----DLAEY--NFLFALSF 1598  
 Db 2062 MADPAVVALASAVOLVSKVITRICRVMDKSTNPQYQLGQMMMDIALIIGRYLIMLSF 2121  
 Qy 1599 TTFPFDEBVAPSGTILATV-----NVALMGLDELKKEEPLQVAGLS 1642  
 Db 2122 NNCI--DVATSVYLFHTITFLVCGSPSKRASTHGLVNIHSLCTINP-----S 2171  
 Qy 1643 TOAEGTRTLK--SLIMFTWENCFYLLI-----SOMRYLRDPAVHPRDK--ORMKOEL 1691  
 Db 2172 FSEBAQVLRSLDEFSLPK--FYLLFGISKVSAVAFAFRSSCHRPDKMLGNERYQPL 2230  
 Qy 1692 SSELSTL--LSGL 1702  
 Db 2231 PADRERLSLPSL 2242  
 RESULT 38  
 T40186  
 Probable phosphatidylinositol 3-kinase - fission yeast (Schizosaccharomyces pombe)  
 C:Species: Schizosaccharomyces pombe  
 C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
 C:Accession: T40186  
 R:Wood, V., Rajandream, M.A.; Barrall, B.G.; Lauber, J.; Hilbert, H.; Duesternoeft, A.  
 Submitted to the EMBL Data Library, February 1998  
 A:Reference number: Z21910  
 A:Accession: T40186  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-2335 <MOO>  
 A:Cross-references: UNIPROT:O14356, EMBL:Z97992, PIDD:CAI10805.1, GSPDB:GNO0067, SPDB:SH

A:Experimental source: strain 972h-, cosmid c30D10  
 C:Genetics:  
 A:Gene: SPDB:SPBC30D10.10c  
 A:Map position: 2  
 C:Superfamily: yeast TOR2 protein  
 Query Match 1.5%; Score 135.5; DB 2; Length 2335;  
 Best Local Similarity 18.2%; Pred. No. 3.7;  
 Matches 279; Conservative 219; Mismatches 526; Indels 507; Gaps 71;  
 Qy 209 VOCLREOSMLL---ELIF-----LYAYFEMAPSDLLVTKMFKEGFGSGRQTRRLVD 259  
 Db 198 VVCCREKAVQLQCFNEVTLQAEHGLRQSSVEYLHSGSLAYKELEPEKSG-----SFR 249  
 Qy 260 ETMDPFDRLGYFSALLVGEQMDIESLHKCALDREHLPQAOGLICOMDCLMLTFGD 319  
 Db 250 EHYTFCD-----LALLRHRD--NSTRICIVLLPFLTSEYNPKKFOQRIYDSFV----- 299  
 Qy 320 IPHAPVLLAMALLRHTLPEETSVVKKIGTAIQLN--VFOYLTRLQSLASGNDCT 377  
 Db 300 -----YLSHRKDKESLAFEAIGRIAMAVNEMIPYLQNILKVIRO----- 342  
 Qy 378 TSTRCM-----CYGGLSFVLTSLHLHTLGNQODIIDTACEVLADP--SLPELFWGTE 428  
 Db 343 TLTAKEKTOYEKPEVEECIGMLAAAVKL-----ELLEDSSRLGLILFSC 388  
 Qy 429 PTSGGLITLDSVCGMPHILSP-----LLOLLRALVSGSKTAKKYSPIDKMSFYNELYKH 484  
 Db 389 LSVHLRQALVKAENIPPLAPIQERILNVSQILTKNEIRI-----NDITYTP 438  
 Qy 485 KPHDVISHEGDTLMRQTPRLVPLGGQTNLRIPQGTVGQVMDLDRAYLVREMSYSWT 544  
 Db 439 SFNTIYASRPDQSKSTESILAL-----ETLG-----TFNFGYS 475  
 Qy 545 LFTCEIMLHV-----VSTADVQHQORVYK-----IIDLVAKYS 581  
 Db 476 LISFQISVLSYENDNSEIRIAARTCCQVARDP--CRKTNPLAVESVAEVLKLT 533  
 Qy 582 TDLSIADCLPITSRIYML--ORLTIVISPPVVIASVCNCLAV----- 624  
 Db 534 -LGIADSDPKIRETVSLDERFDRHLAPDN-----HCLPLANDEVPIREIAIIT 586  
 Qy 625 -LAARNPAKWTDLRHHTGFLPFVAHVSSLSQMSIABEGMNAAGYGNILMNSBOQGEY 681  
 Db 587 IGRLATVPAHVMSLRKT-----IIQLLS--DMEVSG-----NSRQKE--- 623  
 Qy 682 GVITAFRLTLTVKQGLSTQSGVLPVCMFLKMLPBYHMKRVNHSVGRBQIGLIL 741  
 Db 624 -----ESAQLIKLVSKARTIKPY-----IQ 645  
 Qy 742 ELIHAILNLCHETDLHSHTPSLQFLCICSLAYTEAGQVINIMGIVDTIDVMAAQR 801  
 Db 646 SIHVLILPKRADT---SPGVSSAIIISALGELASVEGDMVDVDRGSMKIL--LVNLDQ 700  
 Qy 802 SDGAEQGGQGLIKTVKLAFSVTN-----NVIRLKPSNVVSPLEQALSQH 848  
 Db 701 SSTLKRLASLKCRKLGRSGYVIQPYLDYRPLGALIGILOEBOPIRREVRJLIGVL 760  
 Qy 849 GAHGNNILAVLAKYI-----HKHPALPRLAIQLKLKATVAPSVVACLG 895  
 Db 761 GA-----LDPTTYITTEVSDDLOSSHNNAHGVQISAAOQPSLENTAYMAVATLLIG 812  
 Qy 896 NDAARIDAFRLTQSLKIED--NRIKWILFELTVAVETOPGLIEFLNLEVXDGSGSK 953  
 Db 813 -----ILKDSLSMHHSSVVOAVVHICQMSKSTVFL--PQVPTFLQV--MQSLSSSA 864  
 Qy 954 EFSLGMSCLHVALELIDSOQDQRYWCPRPLHRAALAFHALMODRDSAMLVLRPKF 1013  
 Db 865 EFYEQQLTTLTISI-----IGBNIRDVSDIFNLKVF 896  
 Qy 1014 MENLSPFLFTLPPSTSEPSLIENTCALMKIKICIEIYVVGSLDQSLKDTLKRSIE 1073  
 Db 897 WESTSTILL-----VILE-----LIDAIALALQDEFKFFYLPQILSCMLKARSLD 940



QY 1074 KRFAYMSGYKSLAVHVAETEGSSCTSLLEYOMLVASRMILLIATTHADIMHLDTSVVR 1133  
 DB 941 -----NTSSRSVSYKVLDS-----FVIGSNIEEMHMLPVIT 974  
 QY 1134 ROLFVLDVGTAKLLLVASVNCRLGSMKCTLLIILRQMKRELGSDVETILGPTL 1192  
 DB 975 RSFERDT-----IFLG-----FRKSALKC-----IAQFQSVNFSDSHARSIIHPLVRMG 1019  
 QY 1193 --EGLVADQOOLMEKTKAV-----FSFETVLQ--MKEMKVSDIPQSOVLVAVC--E 1240  
 DB 1020 KSNDELRA--VIMPTLCIVSOLGYDSIFTPMNVKVLVSHKISH--PAYELLVSRLLKGE 1076  
 QY 1241 TLQEEVIALPQDTHSLALGSATEDK--DSMETDCS-----RSRHRQDQGVCLGLHL 1293  
 DB 1077 PLPRDVVYKFKRPSTKTPFS--TQDEVITKLPVDQASLKAWESSQKLTLD----- 1126  
 QY 1294 AKELCEVEDSDGMLQVTRRLPLPTLLTLEVSILRMKQNLHFEATLHLLTLARTQOG 1353  
 DB 1127 -----DWQDWI-----RRISI--ELKESPSALRSCTL----- 1154  
 QY 1354 ATAVAGAITOSICPLLSYV-----QLSTNGTRQTPSASRKSILDAPSMGVRLMSL 1407  
 DB 1155 -----AGIYHPLARDLFNVSFSLCWDDELTESNKNLVKSIELMNAFN-----ISVEI 1202  
 QY 1408 MEOLLKTLRY-----NFLPEALDFGVYHGER-----TLQCLN-----AVPTVQ 1445  
 DB 1203 LQTLINLAIEYEREDHTLPRIKVISAHASKCNVYAKALHTELOFQVETKEVVSISTIE 1262  
 QY 1446 SLAC-----LEBADHTVGFILQLSNFMKE-----WHFPLPOLMR----- 1479  
 DB 1263 SLRTNNHMLQOSDAVG--MLQYTKENHQPFLKETWY--EKLHWMDALAEYEREREGD 1318  
 QY 1480 -DIVNNGYLCQACTSLSHSRKMLQHTLQNKNGGLSPAVAQVORPPSAAAPS----- 1534  
 DB 1319 SSFFINIGKL--RCYVALGMDHLSLQAOKA---WVTSQEHREAIPLAAMAAMGLQGW 1373  
 QY 1535 ---SSQPADTEASEO-----ALHTVOY 1556  
 DB 1374 MLISEYVSANDRPQDKKEFFSAISAVHLGQY 1404  
 RESULT 39  
 T42735  
 TBP-interacting protein TBP120 - rat  
 C:Species: Rattus norvegicus (Norway rat)  
 C>Date: 11-Jan-2000 #sequence revision 11-Jan-2000 #text change 09-Jul-2004  
 C/Accession: T42735  
 R/Yogobawa, S.; Makino, Y.; Yoshida, T.; Kishimoto, T.; Muramatsu, M.; Tamura, T.  
 Biochem. Biophys. Res. Commun. 229, 612-617, 1996  
 A>Title: Molecular cloning of a novel 120-kDa TBP-interacting protein.  
 A/Reference number: 222253; MUID:97127450; PMID:8954946  
 A/Accession: T42735  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: mRNA  
 A/Residues: 1-1230 <YOG>  
 A/Cross-references: UNIPROT:P97536; EMBL:D87671; NID:g1799569; PTDN:BAAL3432.1; PID:g175  
 C/Genetic: C  
 A/Function: C  
 A/Description: considered to participate in transcription regulation through the interact  
 Query Match 1.54; Score 135; DB 2; Length 1230;  
 Best Local Similarity 19.28; Pred. No. 1.4;  
 Matches 269; Conservative 191; Mismatches 468; Indels 470; Gaps 68;  
 QY 255 RLHVDFTMDPFDRIIGYFSAIILVEGMDIESLHKCALDRRELHQPADGLICQDMDCIM 314  
 DB 65 QNLAVKCLGRLVSKYKVEYETIV-----DTLCNMLSDKQLDIDISSIG-----LK 111  
 QY 315 LTFGDIPHHNAFVLLAMALLRHTLNPETSSVVRKIGGTALQANFOYLTRLLQSLASGNN 374  
 DB 112 TVIGELPPAS-----SSGALANVCCKITGLTSAIAKQE 146

QY 375 DCTTSTACMCYGLLSFVLTSLEHTLGNODIIDTACEVLAD--PS-----LPELFMG 427  
 DB 147 D-----VSQLEALDIDAMMLSRQGLLVNHPILITCLQL-----T 185  
 QY 428 EP-----TSGLGILDSVCG-----MEPHLSPL-----LQLLRALVSG 461  
 DB 186 SPRLAVRKRRTIALGHLVMS--CGNIVFVDILHLLSELSKNDMSMTTRTYIQCIATL--S 242  
 QY 462 KSTAKKYSPDK-----MSFYN-----ELYKHKHVDVISHEDGLMKRQTPKLYPLGQ 512  
 DB 243 RQAGRIRGEYLEKTIPLVVKFQVNDDEL--REYCIQAFSS--FVRCKPEKYVP-- 292  
 QY 513 TNLILPGQGVQVMDRAVLVREY-----SYSS-----SW 543  
 DB 293 -----HVTIINILCTKLITDPNYNYDDEDEDENAMADGDDDDGSDDEYSDDDDMSW 347  
 QY 544 TLFTCEIEMLHVSTADVI--OHQORVKPI-----DLVHKVIS----- 581  
 DB 348 KVRBAAKCLDAVAVSTHEMLPEFYKTVSPALISRFKEREENVKADVFMHAYLSLKQTRP 407  
 QY 582 --TDLSTADCLLPITSHIWMLOPLTVI-----SPRYDVYASCNCLTVLAARPA 631  
 DB 408 VQSWLCDDPDAMEQGETPLTLOSQVPIVYALHKQMEKESYKTRQCCFNMLTELNVYLP 467  
 QY 632 KVMPTDLRHTGLP-FVAHVPVSSLSQMTSABGMNAGVGNILMSEQPGEGVTAIFRL 690  
 DB 468 AL-----TOHIFVLVGGITFSLNDKSS-----SNLKIDN-----LSCLYV 503  
 QY 691 ITTLVKQQLSGTOSQGLVPCVMPFYLKEMLPYHMKRYNSGVREQLCLILELHAILML 750  
 DB 504 ILCHNSFOVHFNHVALPVVAVCGD--PPY-----KISEALLVYQQLVKV 549  
 QY 751 CHETDLHSH--TPSLQFLCISLAYTEA-----GQTVINIMIGVDTID 793  
 DB 550 IRPLDQSSPDATFPIYKIDLFCTIKRLKADIDQEVKRAISCGQILICML----- 600  
 QY 794 MVMMAQQRSDGABEGQGLIKTVKLAFS--VNNVNRILRPPSVV-----SPLEQALSQH 848  
 DB 601 -----GDNIGPDLSTNLQIFLERLKNKEITRLTVKVALTLIAGSPKIDIRPV 647  
 QY 849 GAHGNMLIAVLAKYIKHNDPALFRLAIQLKRLATVAPMSVYACLGND--AAIRDAFL 906  
 DB 648 LGEG---VPIIASFL--RKNQAL-----KLGTSALDILIKNYSDSLTAAMIDAVL 694  
 QY 907 TRLOSKI--EDMRKIMLEFLVAVETQGLIEFLINLEVKQSDGSKESFLMSGCLH 964  
 DB 695 DELPPLISESDMHWQMAISFLTLAKYPS-----SLSKISG-----SILN 736  
 QY 965 AVLELIDSOODRWCPPLHRAAIAFLHLMQDRRSAMLVLRKPKFMEV-----LTSP 1020  
 DB 737 ELIGLVS-----PLIQGGLSMLDPPQ-----ALVVTGTNNLGVWDLIRMLTGP 782  
 QY 1021 LFGTLSPSESTSEPSILETALIMKICLRIYVVYKSLDQSLKDTLKKEFSIE----- 1073  
 DB 783 VYSGSTALTHKQSYYSIAKCVAAITRACPRGPAVVGQFIQDVANGSSTBSIRLLALLSL 842  
 QY 1074 -----KRFAYMSGYKSLAVHVAETEGS--SCTSLLEYQMLVSAMR 1112  
 DB 843 GEVGHNIDLSGOLELKSIVLEAFSPSEEVKSAASVAL-----GSISVGNLDEYLPFV----- 895  
 QY 1113 MLIIATTHADIMHMLTDSVVRQLFL-----DVDGTAKALLVPAASV--CLRLGSMCTL 1166  
 DB 896 -----LQETTSQPRQYLLHSLKEILISSASVGLKPEVENIMALLHGHCA-- 943  
 QY 1167 LLLILRQMKRELGS--VDEILGPLETILEGVLQADQOLMEKTKAKYFS-----AFITVLQ 1219  
 DB 944 -----EBGRNVAVACLGKLT-----LIDPETLPLRLKGYIISGSSVARSVV 987  
 QY 1220 MKEMKVSDIPQSOVLVAVC-----ETLQEEVIALPQDTHSLAL--GSATEDKDSMET 1271  
 DB 988 AVFTTISHDQPIDPLKNCIGDFLKTLEDP-----DLNVRVALVTFNSAAHNPSEL-- 1040

QY 1272 DDCSRHRDQDGVCLGLHL-----AKELCEVEDG-----DSWLQVTR-RLPLP 1318  
DB 1041 -----IRDLDBVLPR---HLVNETKREKLEIPEVEMGPRIKIVVDGLOIRKAAFEQCMY 1090  
QY 1319 TLLTLEVLNRMKONLHFTETATL--HLLLTARTQOGATAVAGITQSLPILLSYQL 1376  
DB 1091 TLDSCLDRLDIPEFLNHVEDGLKDHVDIKM-----LTFPLMIVRL 1130  
QY 1377 STNCTACTPSRSKSLDAPSWPGYVYRLSMLEQLOKTLRYNPLPEALDPGVQVQERTIQ 1436  
DB 1131 ST-----LCPSSAVLQRLD-----RLVEPLRATCTTKVKNVSKQEPF-----KODELKR 1174  
QY 1437 CINAARTVQSLACEAD 1454  
DB 1175 --SAMRAVALLTIPPEAF 1190

## RESULT 40

AC2012  
hypothetical protein a111649 (imported) - Nostoc sp. (strain PCC 7120)  
C:Species: Nostoc sp. PCC 7120  
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C:Date: 14-Dec-2001 #sequence revision 14-Dec-2001 #text change 09-Jul-2004  
C:Accession: AC2012  
R:Kaneko, T.; Nakamura, Y.; Molk, C.P.; Kuritz, T.; Saeamto, S.; Matanabe, A.; Iriyuchi, N.; Shimizu, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S. DNA Res. 8, 205-213, 2001  
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena  
A:Reference number: AB1807; MUID:21595285; PMID:11759840  
A:Accession: AC2012  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1570 <KUR>  
A:Cross-references: UNIPROT:Q8YWG4; GS:BA000019; PIDN:BA878015.1; PID:G17135469; GSPDB:C  
A:Experimental source: strain PCC 7120  
C:Gene: a111649  
C:Superfamily: Streptomyces hygroscopicus probable polyketide synthase module 4; 3-oxoac  
homology; [acyl-carrier-protein] S-malonyltransferase homology  
C:Keywords: carrier protein

Query Match 1.5%; Score 135; DB 2; Length 1570;  
Best Local Similarity 19.6%; Pred. No. 2.1;  
Matches 185; Conservative 116; Mismatches 317; Indels 324; Gaps 44;  
QY 321 PHNAPVLLAMLLHNTLNPETSSVVKIGSTALQINVFQVLTLLQSLASGNDCTTST 380  
DB 800 PH--PVLGY--INACLRKQEVGVI-----LPSLRKG----- 828  
QY 381 ACWCYGLSFLVLSL-ELHTLG---NQODIIDTACEVLADPSLP---ELFWGTEPTSG 433  
DB 829 -----FGERATLTLGLTLYLTHGSHINWQSLYPDCCQWVLPYPMQRESYVSESQP-- 881  
QY 434 GIILDSVCGMPHLLSPLLQLLRALVSGSKTAKVYSLDKMSFYNELYKHKPHDIVISHE 493  
DB 882 -----QFOKAL--PASSLNLTLVAGKTE-----QLTQELSNH-----HQ 913  
QY 494 DGTIMRQTPKLLYPLG-GQTNLAIPQCTGVQWMLDRAVLYRWYS-----YSS 542  
DB 914 LSPBAKQFIPQLQLLTKGSKATKIDP-----LSNARYIEWQLSPLTVDNKTSBAER 967  
QY 543 WTLFTCEI---EMLLHVSTADVI-----QHCRVKPILIDVHKVISTDLSIA 587  
DB 968 WLIFSDNOGLGCLAEAVVNDSCILVSSGESYKXSSSHYQ--INF-----HQADPQQLIA 1021  
QY 588 DCLLPITSRIYM-----LQRLTTVISPPV-----DV 614  
DB 1022 DISPEITKVVYLVWELSDSINLNSQPQCYSHLVYVQALAKISSKIAPOLMIGTQAQAV 1081  
QY 615 IAGC--VM-----CLYTLAARNPAKWTID----- 636  
DB 1082 TPSCHQINVAQTPLMGQVIALLEYPOLMGGLIDFGNODVAATAIATMTAKTGEDRVAF 1141

QY 637 ---LRHTGFLPFAHVPVSLQWIS-AEGMNAAGYNLLMNSHQPOGEYGVTTA----- 686  
DB 1142 RDKRIVARIMPLISALPAPQPLISDGSYILITGGLAL-----GLTLAEWLVQ 1190  
QY 687 -FLRLITTVKQOLGST-----QSQG---LVPCVMFLKEMLPBHYKRRYNSHG 731  
DB 1191 GASHLVLTSHQGLINQSEKQCKIRALENGATVAVKVAADVSDYQMSQLFAOIQLSNPK 1250  
QY 732 VRBOIGLILIELHA--LNLCHETDLHSHTSLQFLCICSLATYBAGQTVININGIGV 789  
DB 1251 LR-----GITHAAGVLA-----DCSTAQ-----M 1269  
QY 790 DTIDMVAAPRSDGAEQGGQGLIKTVLAESVTNNVIRLRPNSNVVSPLEQALSQHG 849  
DB 1270 DSEFTFKVPQPKYTGAMNLHQ---LVQDSLDPFVCE-----SSMSALLSROGLHY 1318  
QY 850 AAGNNLIATLAKYIYHNDPALRLAIQ-----LKRATVAPMSVYAC 893  
DB 1319 AANSPFLDGL---VYHQSLGLPGLSINMGPMAGGATQGYEVGLKRMG-IEPLLEPTAA 1374  
QY 894 LGNDAAAIRDAFLTRLQSKIEDMRIRVMILEFIL---VAVETQGLIELFLMLEVQGS 949  
DB 1375 LQVLGGLLSA-----SMQTVAAIINMSAFGKIYAAKGVAFLEALLTQESKQGS 1424  
QY 950 DGSKEPSLGMWSCIHAVLELIDSQODRQYCPPLHRAIYA--FLHALMODRRDSAMLV 1007  
DB 1425 NGENPFGKLSAALHRRPALTLTQVQEV-AQVLGHSGSVPEVEQGFPMGMDSLMSV- 1482  
QY 1008 KTYPKTMENTSPFLGTLSPSETSEBSILETCALIMKICILEIYVYKSLDQSLKDTL 1067  
DB 1483 ---QFHSLEALLAVSLPETLVECPSPISGDVVSYLNR---EVFAMQLDADDSAMESQ 1534  
QY 1068 KKFSEKRFAYMSGYKSLAVHAETGSSCTSLLEYOMLV 1109  
DB 1535 ASVVIENITIAQLBG-----LSTAFETBALMEQRIAEIQLLIS 1570

Search completed: December 30, 2004, 15:14:54  
Job time: 121.789 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 30, 2004, 15:00:06 ; Search time 0.191018 Seconds  
(without alignments)  
2083.096 Million cell updates/sec

Title: US-10-719-385-26

Perfect score: 36

Sequence: 1 AGGPGCV 6

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 150 summaries

Database :

Issued Patents AA: \*  
1: /cgn2\_6/ptodata/1/1aa/5A\_COMB.pep: \*  
2: /cgn2\_6/ptodata/1/1aa/5B\_COMB.pep: \*  
3: /cgn2\_6/ptodata/1/1aa/6A\_COMB.pep: \*  
4: /cgn2\_6/ptodata/1/1aa/6B\_COMB.pep: \*  
5: /cgn2\_6/ptodata/1/1aa/PTCUS\_COMB.pep: \*  
6: /cgn2\_6/ptodata/1/1aa/backfile1.pep: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	32	88.9	15	4	US-09-142-524D-34
2	32	88.9	15	4	US-09-142-524D-35
3	32	88.9	18	2	US-09-017-205-20
4	32	88.9	20	3	US-08-467-023-35
5	32	88.9	20	3	US-08-467-023-36
6	32	88.9	30	3	US-08-467-023-67
7	32	88.9	30	3	US-08-467-023-71
8	32	88.9	32	4	US-09-082-358B-40
9	32	88.9	54	2	US-08-400-159-16
10	32	88.9	54	3	US-08-611-729A-16
11	32	88.9	54	4	US-09-195-524-16
12	32	88.9	60	4	US-08-467-023-62
13	32	88.9	66	1	US-08-450-360-10
14	32	88.9	91	4	US-09-198-452A-726
15	32	88.9	108	4	US-09-513-999C-8111
16	32	88.9	135	4	US-09-252-991A-17866
17	32	88.9	171	4	US-09-252-991A-20639
18	32	88.9	194	4	US-09-252-991A-21199
19	32	88.9	196	2	US-08-481-814A-9
20	32	88.9	198	3	US-09-068-740A-5
21	32	88.9	226	4	US-09-252-991A-27728
22	32	88.9	264	4	US-09-252-991A-20374
23	32	88.9	275	3	US-09-036-987A-7
24	32	88.9	275	3	US-09-370-700-7
25	32	88.9	275	4	US-09-603-207-7
26	32	88.9	290	4	US-09-252-991A-23586
27	32	88.9	365	2	US-08-481-814A-10

28	32	88.9	374	3	US-08-467-023-2	Sequence 2, Appli
29	32	88.9	437	1	US-08-136-119-4	Sequence 4, Appli
30	32	88.9	437	1	US-07-882-711-2	Sequence 2, Appli
31	32	88.9	437	2	US-08-723-415B-13	Sequence 13, Appli
32	32	88.9	437	2	US-08-461-814A-6	Sequence 6, Appli
33	32	88.9	437	2	US-08-462-174-2	Sequence 2, Appli
34	32	88.9	437	3	US-08-801-092-1	Sequence 1, Appli
35	32	88.9	437	3	US-09-189-627A-13	Sequence 13, Appli
36	32	88.9	437	3	US-09-242-737-2	Sequence 2, Appli
37	32	88.9	437	3	US-09-315-113-1	Sequence 1, Appli
38	32	88.9	437	3	US-09-710-861-13	Sequence 13, Appli
39	32	88.9	447	1	US-08-450-360-4	Sequence 4, Appli
40	32	88.9	476	1	US-08-139-937-14	Sequence 14, Appli
41	32	88.9	476	5	PCT-US93-11310-14	Sequence 14, Appli
42	32	88.9	486	1	US-08-450-360-2	Sequence 2, Appli
43	32	88.9	744	2	US-08-462-080B-2	Sequence 2, Appli
44	32	88.9	744	3	US-08-462-090-2	Sequence 2, Appli
45	32	88.9	744	3	US-08-463-461-2	Sequence 2, Appli
46	32	88.9	1010	3	US-08-882-046-7	Sequence 7, Appli
47	32	88.9	1010	4	US-09-556-047-7	Sequence 7, Appli
48	32	88.9	1036	4	US-09-068-740A-6	Sequence 6, Appli
49	32	88.9	1067	4	US-09-579-536C-18	Sequence 18, Appli
50	32	88.9	1160	3	US-08-808-599A-24	Sequence 24, Appli
51	32	88.9	1187	3	US-09-068-740A-7	Sequence 7, Appli
52	32	88.9	1193	2	US-08-400-159-10	Sequence 10, Appli
53	32	88.9	1193	3	US-08-611-729A-10	Sequence 10, Appli
54	32	88.9	1193	4	US-09-195-524-10	Sequence 10, Appli
55	32	88.9	1208	4	US-09-199-865-1	Sequence 1, Appli
56	32	88.9	1218	4	US-08-400-159-6	Sequence 6, Appli
57	32	88.9	1218	3	US-08-611-729A-6	Sequence 6, Appli
58	32	88.9	1218	3	US-08-882-046-2	Sequence 2, Appli
59	32	88.9	1218	3	US-09-214-278-7	Sequence 7, Appli
60	32	88.9	1218	3	US-09-058-740A-11	Sequence 11, Appli
61	32	88.9	1218	4	US-09-855-722-7	Sequence 7, Appli
62	32	88.9	1218	4	US-09-566-047-2	Sequence 2, Appli
63	32	88.9	1218	4	US-09-917-254-85	Sequence 85, Appli
64	32	88.9	1218	4	US-09-195-524-6	Sequence 6, Appli
65	32	88.9	1218	4	US-09-579-536C-1	Sequence 1, Appli
66	32	88.9	1219	3	US-08-882-046-5	Sequence 5, Appli
67	32	88.9	1219	4	US-09-566-047-5	Sequence 5, Appli
68	32	88.9	1404	2	US-08-400-159-2	Sequence 2, Appli
69	32	88.9	1404	3	US-08-611-729A-2	Sequence 2, Appli
70	32	88.9	1404	2	US-09-195-524-2	Sequence 2, Appli
71	31	86.1	27	3	US-09-219-983A-35	Sequence 35, Appli
72	31	86.1	780	4	US-10-148-806-35	Sequence 35, Appli
73	30	83.3	92	4	US-09-513-999C-6101	Sequence 6101, Ap
74	29	80.6	61	4	US-09-621-976-7028	Sequence 7028, Ap
75	29	80.6	67	4	US-09-621-976-6596	Sequence 6596, Ap
76	29	80.6	69	4	US-09-621-976-6022	Sequence 6022, Ap
77	29	80.6	103	4	US-09-513-999C-7000	Sequence 7000, Ap
78	29	80.6	106	3	US-09-199-637A-233	Sequence 233, Ap
79	29	80.6	126	4	US-09-205-288-1076	Sequence 1076, Ap
80	29	80.6	143	4	US-09-252-991A-17861	Sequence 17861, A
81	29	80.6	147	4	US-09-252-991A-31725	Sequence 31725, A
82	29	80.6	235	4	US-09-252-991A-24522	Sequence 24522, A
83	29	80.6	244	4	US-09-769-482-12	Sequence 12, Appli
84	29	80.6	249	4	US-09-311-784A-16	Sequence 16, Appli
85	29	80.6	254	4	US-09-270-767-11730	Sequence 11730, A
86	29	80.6	269	4	US-09-311-021-66	Sequence 66, Appli
87	29	80.6	288	4	US-09-091-952A-4	Sequence 4, Appli
88	29	80.6	306	4	US-09-091-952A-3	Sequence 3, Appli
89	29	80.6	323	4	US-09-252-991A-30172	Sequence 30172, A
90	29	80.6	367	3	US-08-467-023-95	Sequence 95, Appli
91	29	80.6	370	3	US-08-467-023-97	Sequence 97, Appli
92	29	80.6	399	4	US-09-270-767-44122	Sequence 44122, A
93	29	80.6	423	2	US-08-955-713-2	Sequence 2, Appli
94	29	80.6	480	4	US-09-535-909-4	Sequence 4, Appli
95	29	80.6	580	4	US-09-252-991A-29629	Sequence 29629, A
96	29	80.6	592	1	US-09-134-000C-5477	Sequence 5477, Ap
97	29	80.6	968	1	US-08-434-730-14	Sequence 14, Appli
98	29	80.6	2710	2	US-08-568-459A-12	Sequence 12, Appli
99	29	80.6	2710	2	US-08-487-826B-12	Sequence 12, Appli
100	29	80.6	2710	3	US-09-210-288-12	Sequence 12, Appli

101	29	80.6	3060	2	US-08-487-826B-14	Sequence 14, Appl
102	28	77.8	8	1	US-08-079-812-34	Sequence 34, Appl
103	28	77.8	14	2	US-09-741-171-5	Sequence 5, Appl
104	28	77.8	18	2	US-09-017-205-19	Sequence 19, Appl
105	28	77.8	19	4	US-09-741-171-4	Sequence 4, Appl
106	28	77.8	31	4	US-09-084-303B-295	Sequence 295, Appl
107	28	77.8	32	4	US-09-084-303B-285	Sequence 143, Appl
108	28	77.8	34	4	US-09-749-637A-143	Sequence 16, Appl
109	28	77.8	37	1	US-08-149-839B-16	Sequence 16, Appl
110	28	77.8	37	1	US-08-149-839B-18	Sequence 16, Appl
111	28	77.8	37	1	US-08-451-568-16	Sequence 16, Appl
112	28	77.8	37	1	US-08-451-568-18	Sequence 16, Appl
113	28	77.8	37	1	US-08-451-566-16	Sequence 18, Appl
114	28	77.8	37	1	US-08-451-566-18	Sequence 16, Appl
115	28	77.8	37	2	US-08-777-113-16	Sequence 16, Appl
116	28	77.8	37	2	US-08-777-113-18	Sequence 16, Appl
117	28	77.8	37	4	US-09-298-574-16	Sequence 18, Appl
118	28	77.8	37	4	US-09-298-574-18	Sequence 213, Appl
119	28	77.8	40	3	US-08-851-843A-213	Sequence 332, Appl
120	28	77.8	40	3	US-08-851-843A-332	Sequence 213, Appl
121	28	77.8	40	3	US-08-854-050-213	Sequence 213, Appl
122	28	77.8	40	3	US-09-430-333-213	Sequence 332, Appl
123	28	77.8	40	4	US-09-402-181B-332	Sequence 332, Appl
124	28	77.8	40	4	US-09-721-456-332	Sequence 3, Appl
125	28	77.8	40	4	US-09-741-171-3	Sequence 5, Appl
126	28	77.8	46	3	US-09-391-799-5	Sequence 573, Appl
127	28	77.8	54	4	US-09-149-476-573	Sequence 40, Appl
128	28	77.8	58	1	US-08-676-125A-40	Sequence 210, Appl
129	28	77.8	58	1	US-08-086-328C-31	Sequence 42, Appl
130	28	77.8	58	1	US-08-086-328C-210	Sequence 44, Appl
131	28	77.8	58	2	US-09-136-012A-40	Sequence 42, Appl
132	28	77.8	58	2	US-08-676-124-42	Sequence 44, Appl
133	28	77.8	58	3	US-08-676-124-43	Sequence 98, Appl
134	28	77.8	58	3	US-08-676-124-44	Sequence 109, Appl
135	28	77.8	58	3	US-08-676-124-98	Sequence 111, Appl
136	28	77.8	58	3	US-08-676-124-100	Sequence 42, Appl
137	28	77.8	58	3	US-08-676-124-106	Sequence 43, Appl
138	28	77.8	58	3	US-08-676-124-109	Sequence 44, Appl
139	28	77.8	58	3	US-08-676-124-111	Sequence 42, Appl
140	28	77.8	58	3	US-09-414-878-42	Sequence 43, Appl
141	28	77.8	58	3	US-09-414-878-43	Sequence 44, Appl
142	28	77.8	58	3	US-09-414-878-44	Sequence 98, Appl
143	28	77.8	58	3	US-09-414-878-98	Sequence 109, Appl
144	28	77.8	58	3	US-09-414-878-100	Sequence 106, Appl
145	28	77.8	58	3	US-09-414-878-106	Sequence 111, Appl
146	28	77.8	58	3	US-09-414-878-109	Sequence 42, Appl
147	28	77.8	58	3	US-09-414-878-111	Sequence 43, Appl
148	28	77.8	58	3	US-09-240-136-42	Sequence 44, Appl
149	28	77.8	58	3	US-09-240-136-43	Sequence 44, Appl
150	28	77.8	58	3	US-09-240-136-44	Sequence 44, Appl

## ALIGNMENTS

RESULT 1  
US-09-142-524D-34  
Sequence 34, Application US/09142524D  
Patent No. 671976  
GENERAL INFORMATION:  
APPLICANT: Sone, Toshio  
APPLICANT: Kume, Akinori  
APPLICANT: Dairiki, Kazuo  
APPLICANT: Iwama, Akiro  
APPLICANT: Kino, Kohsuke  
TITLE OF INVENTION: Peptide-Based Immunotherapeutic Agent for Treating Allergic Disease  
FILE REFERENCE: SPO-103  
CURRENT APPLICATION NUMBER: US/09/142,524D  
CURRENT FILING DATE: 1998-09-09  
PRIOR APPLICATION NUMBER: PCT/JP97/00740  
PRIOR FILING DATE: 1997-03-10  
NUMBER OF SEQ ID NOS: 174  
SOFTWARE: PatentIn version 3.1

SEQ ID NO 34  
LENGTH: 15  
TYPE: PRT  
ORGANISM: Cryptosporidia japonica  
FEATURE:  
NAME/KEY: MISC FEATURE  
LOCATION: (1)..(15)  
OTHER INFORMATION: Cryj1 peptide, Figure 1, Row 120  
US-09-142-524D-34

Query Match 88.9%; Score 32; DB 4; Length 15;  
Best Local Similarity 100.0%; Pred. No. 47;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GGPCV 6  
DB 9 GGPCV 13

RESULT 2  
US-09-142-524D-35  
Sequence 35, Application US/09142524D  
Patent No. 671976  
GENERAL INFORMATION:  
APPLICANT: Sone, Toshio  
APPLICANT: Kume, Akinori  
APPLICANT: Dairiki, Kazuo  
APPLICANT: Iwama, Akiro  
APPLICANT: Kino, Kohsuke  
TITLE OF INVENTION: Peptide-Based Immunotherapeutic Agent for Treating Allergic Disease  
FILE REFERENCE: SPO-103  
CURRENT APPLICATION NUMBER: US/09/142,524D  
CURRENT FILING DATE: 1998-09-09  
PRIOR APPLICATION NUMBER: PCT/JP97/00740  
PRIOR FILING DATE: 1997-03-10  
NUMBER OF SEQ ID NOS: 174  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 35  
LENGTH: 15  
TYPE: PRT  
ORGANISM: Cryptosporidia japonica  
FEATURE:  
NAME/KEY: MISC FEATURE  
LOCATION: (1)..(15)  
OTHER INFORMATION: Cryj1 peptide, Figure 1, Row 21  
US-09-142-524D-35

Query Match 88.9%; Score 32; DB 4; Length 15;  
Best Local Similarity 100.0%; Pred. No. 47;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GGPCV 6  
DB 4 GGPCV 8

RESULT 3  
US-09-017-205-20  
Sequence 20, Application US/09017205  
Patent No. 5965357  
GENERAL INFORMATION:  
APPLICANT: Marsden, Howard S  
TITLE OF INVENTION: PEPTIDE STRUCTURES AND THEIR USE IN  
DIAGNOSIS OF HERPES SIMPLEX VIRUS TYPE 2  
NUMBER OF SEQUENCES: 86  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Nixon & Vandierke PC  
STREET: 8th Floor, 1100 No. 5965357th Glebe Road  
CITY: Arlington  
STATE: Virginia  
COUNTRY: USA  
ZIP: 22201-4714  
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/017,205  
FILING DATE: 02-FEB-1998  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Mitchard, Leonard C  
REGISTRATION NUMBER: 29,009  
REFERENCE/DOCKET NUMBER: 604-436  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)816-4000  
TELEFAX: (703)816-4100  
INFORMATION FOR SEQ ID NO: 20:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 18 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide from HSV-2 glycoprotein G  
FRAGMENT TYPE: internal  
US-09-017-205-20

Query Match 88.9%; Score 32; DB 2; Length 18;  
Best Local Similarity 100.0%; Pred. No. 55;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GPCV 6  
|||||  
Db 5 GPCV 9

RESULT 4  
US-08-467-023-35  
Sequence 35, Application US/08467023  
Patent No. 6090386  
GENERAL INFORMATION:  
APPLICANT: Griffith, Irwin J.;  
APPLICANT: Pollock, Joanne;  
APPLICANT: Bond, Julian F.;  
APPLICANT: Garman, Richard D;  
APPLICANT: Kuo, Mei-Chang;  
APPLICANT: Yeung, Siu-mei H.;  
APPLICANT: Brauer, Andrew;  
APPLICANT: Exley, Mark A.;  
APPLICANT: Powers, Steven P.  
TITLE OF INVENTION: Allergenic Proteins And Peptides From  
TITLE OF INVENTION: Japanese Cedar Pollen  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/467,023  
FILING DATE: June 6, 1995  
CLASSIFICATION: 424  
INFORMATION FOR SEQ ID NO: 20:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 20 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: internal  
US-08-467-023-36

TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 227-7400  
TELEFAX: (617) 227-5941  
INFORMATION FOR SEQ ID NO: 35:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 20 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: internal  
US-08-467-023-35

Query Match 88.9%; Score 32; DB 3; Length 20;  
Best Local Similarity 100.0%; Pred. No. 60;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GPCV 6  
|||||  
Db 14 GPCV 18

RESULT 5  
US-08-467-023-36  
Sequence 36, Application US/08467023  
Patent No. 6090386  
GENERAL INFORMATION:  
APPLICANT: Griffith, Irwin J.;  
APPLICANT: Pollock, Joanne;  
APPLICANT: Bond, Julian F.;  
APPLICANT: Garman, Richard D;  
APPLICANT: Kuo, Mei-Chang;  
APPLICANT: Yeung, Siu-mei H.;  
APPLICANT: Brauer, Andrew;  
APPLICANT: Exley, Mark A.;  
APPLICANT: Powers, Steven P.  
TITLE OF INVENTION: Allergenic Proteins And Peptides From  
TITLE OF INVENTION: Japanese Cedar Pollen  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/467,023  
FILING DATE: June 6, 1995  
CLASSIFICATION: 424  
INFORMATION FOR SEQ ID NO: 36:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 20 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: internal  
US-08-467-023-36

Query Match 88.9%; Score 32; DB 3; Length 20;  
Best Local Similarity 100.0%; Pred. No. 60;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GPCV 6  
|||||  
Db 4 GPCV 8

## RESULT 6

US-08-467-023-67  
; Sequence 67, Application US/08467023  
; Patent No. 6090386  
; GENERAL INFORMATION:  
; APPLICANT: Griffee, Irwin J.;  
; APPLICANT: Pollock, Joanne;  
; APPLICANT: Bond, Julian F.;  
; APPLICANT: Garman, Richard D;  
; APPLICANT: Kuo, Mei-Chang;  
; APPLICANT: Yeung, Siu-mei H.;  
; APPLICANT: Brauer, Andrew;  
; APPLICANT: Exley, Mark A.;  
; APPLICANT: Powers, Steven P.  
; TITLE OF INVENTION: Allergenic Proteins And Peptides From  
; TITLE OF INVENTION: Japanese Cedar Pollen  
; NUMBER OF SEQUENCES: 261  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Immunologic Pharmaceutical Corporation, Inc.  
; STREET: 610 Lincoln St  
; CITY: Waltham  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02154  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/467, 023  
; FILING DATE: June 6, 1995  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/350,225  
; FILING DATE: December 6, 1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Jane E. Remillard  
; REGISTRATION NUMBER: 38,872  
; REFERENCE/DOCKET NUMBER: 025.6 USD2 (IMI-028CPD2)  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 227-7400  
; TELEFAX: (617) 227-5941  
; INFORMATION FOR SEQ ID NO: 67:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 30 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; FRAGMENT TYPE: internal  
; US-08-467-023-67

Query Match 88.9%; Score 32; DB 3; Length 30;  
Best Local Similarity 100.0%; Pred. No. 85;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GPCV 6  
|||||  
Db 24 GPCV 28

## RESULT 7

US-08-467-023-71  
; Sequence 71, Application US/08467023

; Patent No. 6090386  
; GENERAL INFORMATION:  
; APPLICANT: Griffee, Irwin J.;  
; APPLICANT: Pollock, Joanne;  
; APPLICANT: Bond, Julian F.;  
; APPLICANT: Garman, Richard D;  
; APPLICANT: Kuo, Mei-Chang;  
; APPLICANT: Yeung, Siu-mei H.;  
; APPLICANT: Brauer, Andrew;  
; APPLICANT: Exley, Mark A.;  
; APPLICANT: Powers, Steven P.  
; TITLE OF INVENTION: Allergenic Proteins And Peptides From  
; TITLE OF INVENTION: Japanese Cedar Pollen  
; NUMBER OF SEQUENCES: 261  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Immunologic Pharmaceutical Corporation, Inc.  
; STREET: 610 Lincoln St  
; CITY: Waltham  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02154  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/467, 023  
; FILING DATE: June 6, 1995  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/350,225  
; FILING DATE: December 6, 1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Jane E. Remillard  
; REGISTRATION NUMBER: 38,872  
; REFERENCE/DOCKET NUMBER: 025.6 USD2 (IMI-028CPD2)  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 227-7400  
; TELEFAX: (617) 227-5941  
; INFORMATION FOR SEQ ID NO: 71:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 30 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
; FRAGMENT TYPE: internal  
; US-08-467-023-71

Query Match 88.9%; Score 32; DB 3; Length 30;  
Best Local Similarity 100.0%; Pred. No. 85;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GPCV 6  
|||||  
Db 24 GPCV 28

## RESULT 8

US-09-082-358B-40  
; Sequence 40, Application US/09082358B  
; Patent No. 6469153  
; GENERAL INFORMATION:  
; APPLICANT: Goff, Stephen P.  
; APPLICANT: Li, Xingqiang  
; TITLE OF INVENTION: EIP-1, EIP-3 GENES, ENVELOPE-INTERACTING PROTEINS,  
; FILE REFERENCE: 0575/54804  
; CURRENT APPLICATION NUMBER: US/09/082,358B  
; CURRENT FILING DATE: 1998-05-20  
; NUMBER OF SEQ ID NOS: 106  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 40

LENGTH: 32  
TYPE: PRT  
ORGANISM: murine  
US-09-082-358B-40

Query Match 88.9%; Score 32; DB 4; Length 32;  
Best Local Similarity 100.0%; Pred. No. 90;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGPC 5  
Db 17 AGGPC 21

RESULT 9  
US-08-400-159-16  
Sequence 16, Application US/08400159  
Patent No. 5869282  
GENERAL INFORMATION:  
APPLICANT: Ish-Horowitz, David  
APPLICANT: Henrique, Domingos M.P.  
APPLICANT: Lewis, Julian H.  
APPLICANT: Myat, Anna M.  
APPLICANT: Fleming, Robert J.  
APPLICANT: Artavanis-Tsakonas, Spyridon  
APPLICANT: Mann, Robert S.  
TITLE OF INVENTION: NUCLEOTIDE AND PROTEIN SEQUENCES OF THE  
TITLE OF INVENTION: SERRATE GENE AND METHODS BASED THEREON  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/400.159  
FILING DATE: 07-MAR-1995  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 7326-029  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741/8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 54 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-400-159-16

Query Match 88.9%; Score 32; DB 2; Length 54;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGPC 5  
Db 5 AGGPC 9

RESULT 10  
US-08-611-729A-16

Sequence 16, Application US/08611729A  
Patent No. 6004924  
GENERAL INFORMATION:

APPLICANT: Ish-Horowitz, David  
APPLICANT: Henrique, Domingos M.P.  
APPLICANT: Lewis, Julian H.  
APPLICANT: Myat, Anna M.  
APPLICANT: Fleming, Robert J.  
APPLICANT: Artavanis-Tsakonas, Spyridon  
APPLICANT: Mann, Robert S.  
TITLE OF INVENTION: NUCLEOTIDE AND PROTEIN SEQUENCES OF THE  
TITLE OF INVENTION: SERRATE GENE AND METHODS BASED THEREON  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/611.729A  
FILING DATE: 06-MAR-1996  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 7326-037  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741/8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 54 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-611-729A-16

Query Match 88.9%; Score 32; DB 3; Length 54;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGPC 5  
Db 5 AGGPC 9

RESULT 11  
US-08-195-524-16  
Sequence 16, Application US/09195524  
Patent No. 6703489  
GENERAL INFORMATION:  
APPLICANT: Ish-Horowitz, David  
APPLICANT: Henrique, Domingos M.P.  
APPLICANT: Lewis, Julian H.  
APPLICANT: Myat, Anna M.  
APPLICANT: Fleming, Robert J.  
APPLICANT: Artavanis-Tsakonas, Spyridon  
APPLICANT: Mann, Robert S.  
TITLE OF INVENTION: NUCLEOTIDE AND PROTEIN SEQUENCES OF THE  
TITLE OF INVENTION: SERRATE GENE AND METHODS BASED THEREON  
NUMBER OF SEQUENCES: 20  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds

STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/195,524  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/611,729  
FILING DATE: 06-MAR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 7326-037  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-9741/8864  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 54 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-09-195-524-16

Query Match 88.9%; Score 32; DB 4; Length 54;  
Best Local Similarity 100.0%; Pred. No. 1.4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGPC 5  
|||||  
Db 5 AGPC 9

RESULT 12  
US-08-467-023-62  
Sequence 62, Application US/08467023  
Patent No. 6090386  
GENERAL INFORMATION:  
APPLICANT: Griffith, Irwin J.;  
APPLICANT: Pollock, Joanne;  
APPLICANT: Bond, Julian F.;  
APPLICANT: Garman, Richard D;  
APPLICANT: Kuo, Mei-Chang;  
APPLICANT: Yeung, Siu-mei H.;  
APPLICANT: Brauer, Andrew;  
APPLICANT: Ekley, Mark A.;  
APPLICANT: Powers, Steven P.  
TITLE OF INVENTION: Allergenic Proteins And Peptides From  
NUMBER OF SEQUENCES: 261  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Immunologic Pharmaceutical Corporation, Inc.  
STREET: 610 Lincoln St  
CITY: Walcham  
STATE: MA  
COUNTRY: USA  
ZIP: 02154  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/467,023

FILING DATE: June 6, 1995  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/350,225  
FILING DATE: December 6, 1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Jane E. Remillard  
REGISTRATION NUMBER: 38,872  
REFERENCE/DOCKET NUMBER: 025.6 USD2 (IMI-028CPD2)  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 227-7400  
TELEFAX: (617) 227-5941  
INFORMATION FOR SEQ ID NO: 62:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 60 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
FRAGMENT TYPE: internal  
US-08-467-023-62

Query Match 88.9%; Score 32; DB 3; Length 60;  
Best Local Similarity 100.0%; Pred. No. 1.5e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GPCV 6  
|||||  
Db 44 GPCV 48

RESULT 13  
US-08-450-360-10  
Sequence 10, Application US/08450360  
Patent No. 5656457  
GENERAL INFORMATION:  
APPLICANT: Parkes, Deborah Lynn  
APPLICANT: Coates, Stephen Ralph  
TITLE OF INVENTION: Herpes Simplex Virus Type 2-Glycoprotein G  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Leona L. Lauder  
STREET: 6 Mariposa Court  
CITY: Tiburon  
STATE: California  
COUNTRY: USA  
ZIP: 94920  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/450,360  
FILING DATE: 25-MAY-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/129,021  
FILING DATE: 29-SEP-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Lauder, Leona L.  
REGISTRATION NUMBER: 30,863  
REFERENCE/DOCKET NUMBER: D-0012.2A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-435-2034  
TELEFAX: 415-435-0727  
GENERAL INFORMATION:  
INFORMATION FOR SEQ ID NO: 10:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 66 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide



US-08-450-360-10

## Query Match

88.9%; Score 32; DB 1; Length 66;  
Best Local Similarity 100.0%; Pred. No. 1.7e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;QY 2 AGPCV 6  
|||||  
DB 18 AGPCV 22RESULT 14  
US-09-198-452A-726

Sequence 726, Application US/09198452A  
Patent No. 6559284  
GENERAL INFORMATION:  
APPLICANT: Griffiths, R.  
TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments thereof and uses thereof, in particular for the diagnosis, prevention and treatment of infection  
FILE REFERENCE: 9710-003-999  
CURRENT APPLICATION NUMBER: US/09/198,452A  
CURRENT FILING DATE: 1998-11-24  
NUMBER OF SEQ ID NOS: 6849  
SEQ ID NO 726  
LENGTH: 91  
TYPE: PRT  
ORGANISM: Chlamydia pneumoniae  
US-09-198-452A-726

## Query Match

88.9%; Score 32; DB 4; Length 91;  
Best Local Similarity 100.0%; Pred. No. 2.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;QY 1 AGPC 5  
|||||  
DB 78 AGPC 82

RESULT 15  
US-09-513-999C-8111  
Sequence 811, Application US/09513999C  
Patent No. 6783961  
GENERAL INFORMATION:  
APPLICANT: Dumas Mline Edwards, J.B.  
APPLICANT: Duclet, A.  
TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.  
Patent No. 6783961  
FILE REFERENCE: 59, US2, REG  
CURRENT APPLICATION NUMBER: US/09/513,999C  
CURRENT FILING DATE: 2000-02-24  
PRIOR APPLICATION NUMBER: US 60/122,487  
PRIOR FILING DATE: 1999-02-26  
NUMBER OF SEQ ID NOS: 36681  
SOFTWARE: Patent.pm  
SEQ ID NO 8111  
LENGTH: 108  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: 6  
OTHER INFORMATION: Xaa=Met or Arg  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: 23  
OTHER INFORMATION: Xaa=Leu or Val  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: 89  
OTHER INFORMATION: Xaa=Gly or Ser  
US-09-513-999C-8111

## Query Match

88.9%; Score 32; DB 4; Length 108;  
Best Local Similarity 100.0%; Pred. No. 2.5e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;QY 1 AGPC 5  
|||||  
DB 24 AGPC 28RESULT 16  
US-09-252-991A-17866

Sequence 17866, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 17866  
LENGTH: 135  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-17866

## Query Match

88.9%; Score 32; DB 4; Length 135;  
Best Local Similarity 100.0%; Pred. No. 3.1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;QY 1 AGPC 5  
|||||  
DB 72 AGPC 76

RESULT 17  
US-09-252-991A-20639  
Sequence 20639, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al.  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 20639  
LENGTH: 171  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-20639

Query Match  
Best Local Similarity 100.0%; Pred. No. 3.7e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGPC 5  
|||||  
DB 115 AGPC 119

RESULT 18

US-09-252-991A-21199  
; Sequence 21199, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 21199  
; LENGTH: 194  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-21199

Query Match 88.9%; Score 32; DB 4; Length 194;  
Best Local Similarity 100.0%; Pred. No. 4.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GGPCV 6  
14 GGPCV 18

RESULT 19  
US-08-481-814A-9  
; Sequence 9, Application US/08481814A  
; Patent No. 5869040  
; GENERAL INFORMATION:  
; APPLICANT: Heu, Yen-Ming  
; TITLE OF INVENTION: GENE THERAPY METHODS AND COMPOSITIONS  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Biogen, Inc.  
; STREET: 14 Cambridge Center  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: United States of America  
; ZIP: 02142

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/481,814A  
; FILING DATE:  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kaplan, Warren A  
; REFERENCE/DOCKET NUMBER: A001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617-679-2000  
; TELEFAX: 617-679-2838  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 196 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: Protein  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
US-08-481-814A-9

Query Match 88.9%; Score 32; DB 2; Length 196;  
Best Local Similarity 100.0%; Pred. No. 4.2e+02;

Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 AGGPC 5  
8 AGGPC 12

RESULT 20  
US-09-068-740A-5  
; Sequence 5, Application US/09068740A  
; Patent No. 6337387  
; GENERAL INFORMATION:  
; APPLICANT: SAKANO, SEIJI  
; TITLE OF INVENTION: DIFFERENTIATION-SUPPRESSIVE POLYPEPTIDE  
; FILE REFERENCE: KP-8447  
; CURRENT APPLICATION NUMBER: US/09/068,740A  
; CURRENT FILING DATE: 1998-06-18  
; PRIOR APPLICATION NUMBER: JP 7-239611  
; PRIOR FILING DATE: 1995-11-17  
; PRIOR APPLICATION NUMBER: JP 7-311811  
; PRIOR FILING DATE: 1995-11-30  
; PRIOR APPLICATION NUMBER: PCT/JP96/03356  
; PRIOR FILING DATE: 1996-11-15  
; NUMBER OF SEQ ID NOS: 48  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 5  
; LENGTH: 198  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-068-740A-5

Query Match 88.9%; Score 32; DB 3; Length 198;  
Best Local Similarity 100.0%; Pred. No. 4.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGGPC 5  
57 AGGPC 61

RESULT 21  
US-09-252-991A-27728  
; Sequence 27728, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 27728  
; LENGTH: 226  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-27728

Query Match 88.9%; Score 32; DB 4; Length 226;  
Best Local Similarity 100.0%; Pred. No. 4.7e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGGPC 5  
104 AGGPC 108

RESULT 22

US-09-252-991A-20374  
; Sequence 20374, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 20374  
; LENGTH: 264  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-20374

Query Match  
Best Local Similarity 88.9%; Score 32; DB 4; Length 264;  
100.0%; Pred. No. 5.4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 AGPC 5  
Db 198 AGPC 202

RESULT 23  
US-09-036-987A-7  
; Sequence 7, Application US/09036987A  
; Patent No. 6143526  
; GENERAL INFORMATION:  
; APPLICANT: Baltz, Richard H.  
; APPLICANT: Broughton, Mary C.  
; APPLICANT: Crawford, Kathryn P.  
; APPLICANT: Madduri, Krishnamurthy  
; APPLICANT: Merlo, Donald J.  
; APPLICANT: Treadway, Patli J.  
; APPLICANT: Turner, Jan R.  
; APPLICANT: Waldron, Clive  
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide  
; TITLE OF INVENTION: Production  
; NUMBER OF SEQUENCES: 39  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dow Agrosciences LLC Patent Department  
; STREET: 9330 Zionsville Road  
; CITY: Indianapolis  
; STATE: Indiana  
; COUNTRY: USA  
; ZIP: 46268  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/036,987A  
; FILING DATE: 09-MAR-1998  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stuart, Donald R  
; REGISTRATION NUMBER: 28,479  
; REFERENCE/DOCKET NUMBER: 50,608  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (317)337-4816  
; TELEFAX: (317)337-4847  
; INFORMATION FOR SEQ ID NO: 7:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 275 amino acids  
; TYPE: amino acid

TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-036-987A-7

Query Match  
Best Local Similarity 88.9%; Score 32; DB 3; Length 275;  
100.0%; Pred. No. 5.6e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 AGPC 5  
Db 27 AGPC 31

RESULT 24  
US-09-370-700-7  
; Sequence 7, Application US/09370700  
; Patent No. 6274350  
; GENERAL INFORMATION:  
; APPLICANT: Baltz, Richard H.  
; APPLICANT: Broughton, Mary C.  
; APPLICANT: Crawford, Kathryn P.  
; APPLICANT: Madduri, Krishnamurthy  
; APPLICANT: Treadway, Patli J.  
; APPLICANT: Turner, Jan R.  
; APPLICANT: Waldron, Clive  
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide  
; FILE REFERENCE: 50489.DIV1  
; CURRENT APPLICATION NUMBER: US/09/370,700  
; CURRENT FILING DATE: 1999-08-09  
; EARLIER APPLICATION NUMBER: US 09/36987  
; EARLIER FILING DATE: 1998-03-09  
; NUMBER OF SEQ ID NOS: 39  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 7  
; LENGTH: 275  
; TYPE: PRT  
; ORGANISM: Saccharopolyspora spinosa  
US-09-370-700-7

Query Match  
Best Local Similarity 88.9%; Score 32; DB 3; Length 275;  
100.0%; Pred. No. 5.6e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 AGPC 5  
Db 27 AGPC 31

RESULT 25  
US-09-603-207-7  
; Sequence 7, Application US/09603207B  
; Patent No. 6521406  
; GENERAL INFORMATION:  
; APPLICANT: Baltz, Richard H.  
; APPLICANT: Broughton, Mary C.  
; APPLICANT: Crawford, Kathryn P.  
; APPLICANT: Madduri, Krishnamurthy  
; APPLICANT: Treadway, Patli J.  
; APPLICANT: Turner, Jan R.  
; APPLICANT: Waldron, Clive  
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide  
; FILE REFERENCE: 50489.DIV1  
; CURRENT APPLICATION NUMBER: US/09/603,207B  
; CURRENT FILING DATE: 2000-06-23  
; EARLIER APPLICATION NUMBER: 09/370,700  
; EARLIER FILING DATE: 1998-03-09  
; NUMBER OF SEQ ID NOS: 39  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 7  
; LENGTH: 275  
; TYPE: PRT  
; ORGANISM: Saccharopolyspora spinosa  
US-09-603-207-7

Query Match 88.9%; Score 32; DB 4; Length 275;  
Best Local Similarity 100.0%; Pred. No. 5.6e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGGPC 5  
11111  
DB 27 AGGPC 31

RESULT 26  
US-09-252-991A-23586  
; Sequence 23586, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 23586  
; LENGTH: 290  
; TYPE: PRF  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-23586

Query Match 88.9%; Score 32; DB 4; Length 290;  
Best Local Similarity 100.0%; Pred. No. 5.9e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGGPC 5  
11111  
DB 18 AGGPC 22

RESULT 27  
US-08-481-814A-10  
; Sequence 10, Application US/08481814A  
; Patent No. 5869040  
; GENERAL INFORMATION:  
; APPLICANT: Hsu, Yen-Ming  
; TITLE OF INVENTION: GENE THERAPY METHODS AND COMPOSITIONS  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Biogen, Inc.  
; STREET: 14 Cambridge Center  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: United States of America  
; ZIP: 02142  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/481,814A  
; FILING DATE:  
; CLASSIFICATION: 424  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kaplan, Warren A  
; REFERENCE/DOCKET NUMBER: A001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617-679-2000  
; TELEFAX: 617-679-2838  
; INFORMATION FOR SEQ ID NO: 10:  
; SEQUENCE CHARACTERISTICS:

LENGTH: 365 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
US-08-481-814A-10

Query Match 88.9%; Score 32; DB 2; Length 365;  
Best Local Similarity 100.0%; Pred. No. 7.1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGGPC 5  
11111  
DB 8 AGGPC 12

RESULT 28  
US-08-467-023-2  
; Sequence 2, Application US/08467023  
; Patent No. 6090386  
; GENERAL INFORMATION:  
; APPLICANT: Griffler, Irwin J.;  
; APPLICANT: Pollock, Joanne;  
; APPLICANT: Bond, Julian F.;  
; APPLICANT: Garman, Richard D;  
; APPLICANT: Kuo, Mei-Chang;  
; APPLICANT: Yeung, Siu-mei H.;  
; APPLICANT: Brauer, Andrew;  
; APPLICANT: Exley, Mark A.;  
; APPLICANT: Powers, Steven P.  
; TITLE OF INVENTION: Allergenic Proteins And Peptides From  
; TITLE OF INVENTION: Japanese Cedar Pollen  
; NUMBER OF SEQUENCES: 261  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Immunologic Pharmaceutical Corporation, Inc.  
; STREET: 610 Lincoln St  
; CITY: Waltham  
; STATE: MA  
; COUNTRY: USA  
; ZIP: 02154

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/467,023  
; FILING DATE: June 6, 1995  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/350,225  
; FILING DATE: December 6, 1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Jane E. Remillard  
; REGISTRATION NUMBER: 38,872  
; REFERENCE/DOCKET NUMBER: 025.6 USD2 (IMI-028CPD2)  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 227-7400  
; TELEFAX: (617) 227-5941  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 374 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-467-023-2

Query Match 88.9%; Score 32; DB 3; Length 374;  
Best Local Similarity 100.0%; Pred. No. 7.3e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGPCV 6  
|||  
Db 125 GGPCV 129

## RESULT 29

US-08-136-119-4  
Sequence 4, Application US/08136119  
Patent No. 5473056  
GENERAL INFORMATION:  
APPLICANT: Heimbrock, David C.  
APPLICANT: Hoyle, Mona I.  
APPLICANT: Olfie, Allen I.  
TITLE OF INVENTION: E 2F-2, A NOVEL MAMMALIAN TRANSCRIPTION  
TITLE OF INVENTION: FACTOR  
NUMBER OF SEQUENCES: 29  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: David A. Muthard  
STREET: P.O. Box 2000, 126 Lincoln Avenue  
CITY: Rahway  
STATE: N.J.  
COUNTRY: USA  
ZIP: 07065-0907  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/136,119  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Muthard, David A.  
REGISTRATION NUMBER: 35,297  
REFERENCE/DOCKET NUMBER: 19042  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (908)594-4720  
TELEFAX: (908)594-4720  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 437 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-136-119-4  
Query Match 88.9%; Score 32; DB 1; Length 437;  
Best Local Similarity 100.0%; Pred. No. 8.3e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 AGPC 5  
|||  
Db 8 AGPC 12

## RESULT 30

US-07-882-711-2  
Sequence 2, Application US/07882711  
Patent No. 5759803  
GENERAL INFORMATION:  
APPLICANT: Kaelin Jr., William G.  
APPLICANT: Flemington, Erik  
APPLICANT: Sellers, William  
APPLICANT: Decaprio, James A.  
APPLICANT: Livingston, David M.  
TITLE OF INVENTION: Retinoblastoma Associated Protein-1 cDNA  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Choate, Hall & Stewart  
STREET: Exchange Place, 53 State Street  
CITY: Boston

STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/882,711  
FILING DATE: 19920513  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Kennedy, Bill  
REGISTRATION NUMBER: 33,407  
REFERENCE/DOCKET NUMBER: DPCI#236  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 227-5020  
TELEFAX: (617) 227-7566  
TELEX: 289374  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 437 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHEICAL: NO  
US-07-882-711-2

Query Match 88.9%; Score 32; DB 1; Length 437;  
Best Local Similarity 100.0%; Pred. No. 8.3e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 AGPC 5  
|||  
Db 8 AGPC 12

## RESULT 31

US-08-723-415B-13  
Sequence 13, Application US/08723415B  
Patent No. 5859199  
GENERAL INFORMATION:  
APPLICANT: Lathangue, Nicholas B.  
APPLICANT: delaluna, Susana  
TITLE OF INVENTION: TRANSCRIPTION FACTOR DP-3 AND ISOFORMS  
TITLE OF INVENTION: THEREOF  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: NIXON & VANDERHYE P.C.  
STREET: 1100 No. 5859199th Gleebe Rd. 8th floor  
CITY: Arlington  
STATE: VA  
COUNTRY: USA  
ZIP: 22201-4741  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/723,415B  
FILING DATE: 30-SEP-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9610195.1  
FILING DATE: 15-MAY-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Crawford, Arthur R.  
REGISTRATION NUMBER: 25,327  
REFERENCE/DOCKET NUMBER: 117-220  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-816-4000

TELEFAX: 703-816-4100  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 437 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-723-415B-13

Query Match 88.9%; Score 32; DB 2; Length 437;  
Best Local Similarity 100.0%; Pred. No. 8.3e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGPG 5  
DB 8 AGGPG 12

## RESULT 32

US-08-481-814A-6  
Sequence 6, Application US/08481814A  
Patent No. 5869040  
GENERAL INFORMATION:  
APPLICANT: Heu, Yen-Ming  
TITLE OF INVENTION: GENE THERAPY METHODS AND COMPOSITIONS  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Biogen, Inc.  
STREET: 14 Cambridge Center  
CITY: Cambridge  
STATE: Massachusetts  
COUNTRY: United States of America  
ZIP: 02142  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/481,814A  
FILING DATE:  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Kaplan, Warren A  
REFERENCE/DOCKET NUMBER: A001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-679-2000  
TELEFAX: 617-679-2838  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 437 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: Protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
US-08-481-814A-6

Query Match 88.9%; Score 32; DB 2; Length 437;  
Best Local Similarity 100.0%; Pred. No. 8.3e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGPG 5  
DB 8 AGGPG 12

RESULT 33  
US-08-462-174-2  
Sequence 2, Application US/08462174  
Patent No. 5981723  
GENERAL INFORMATION:

APPLICANT: Kaelin Jr., William G.  
APPLICANT: Flemington, Erik  
APPLICANT: Sellers, William  
APPLICANT: Decaprio, James A.  
APPLICANT: Livingston, David M.  
TITLE OF INVENTION: Retinoblastoma Associated Protein-1 cDNA  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Choate, Hall & Stewart  
STREET: Exchange Place, 53 State Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02109

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/462,174  
FILING DATE: 05-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/882,711  
FILING DATE: 13-MAY-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Kennedy, Bill  
REGISTRATION NUMBER: 33,407  
REFERENCE/DOCKET NUMBER: DPCI236  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 227-5020  
TELEFAX: (617) 227-7566  
TELEX: 289374  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 437 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
US-08-462-174-2

Query Match 88.9%; Score 32; DB 2; Length 437;  
Best Local Similarity 100.0%; Pred. No. 8.3e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGPG 5  
DB 8 AGGPG 12

## RESULT 34

US-08-801-092-1  
Sequence 1, Application US/08801092  
Patent No. 6074850  
GENERAL INFORMATION:  
APPLICANT: Antelman, Douglas  
APPLICANT: Gregory, Richard J.  
APPLICANT: Wils, Kenneth N.  
TITLE OF INVENTION: Tissue Specific Expression of  
NUMBER OF SEQUENCES: 46  
RETINOBLASTOMA PROTEIN  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP  
STREET: Two Embarcadero Center, 8th Floor  
CITY: San Francisco  
STATE: CA  
COUNTRY: USA  
ZIP: 94111

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/801,092  
FILING DATE: 14-FEB-1997  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/751,517  
FILING DATE: 15-NOV-1996  
CLASSIFICATION: 514  
ATTORNEY/AGENT INFORMATION:  
NAME: Fites, Renee A.  
REGISTRATION NUMBER: 35,136  
REFERENCE/DOCKET NUMBER: 016930-001020  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-576-0200  
TELEFAX: 703-576-0300  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 437 amino acids  
TYPE: amino acid  
STRANDEDNESS: not relevant  
TOPOLOGY: not relevant  
MOLECULE TYPE: peptide  
US-08-801-092-1

Query Match 88.9%; Score 32; DB 3; Length 437;  
Best Local Similarity 100.0%; Pred. No. 8.3e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGPC 5  
Db 8 AGGPC 12

RESULT 35  
US-09-189-627A-13  
Sequence 13, Application US/09189627A  
Patent No. 6159691  
GENERAL INFORMATION:  
APPLICANT: Le Thangue, Nicholas  
TITLE OF INVENTION: TRANSCRIPTION FACTOR DP-3 AND ISOFORMS THEREOF  
FILE REFERENCE: 620-54  
CURRENT APPLICATION NUMBER: US/09/189,627A  
CURRENT FILING DATE: 1998-11-10  
PRIOR APPLICATION NUMBER: 08/723,415  
PRIOR FILING DATE: 1996-09-30  
PRIOR APPLICATION NUMBER: GB 9610195  
PRIOR FILING DATE: 1996-05-15  
NUMBER OF SEQ ID NOS: 25  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 13  
LENGTH: 437  
TYPE: PRT  
ORGANISM: human  
US-09-189-627A-13

Query Match 88.9%; Score 32; DB 3; Length 437;  
Best Local Similarity 100.0%; Pred. No. 8.3e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGPC 5  
Db 8 AGGPC 12

RESULT 36  
US-09-242-737-2  
Sequence 2, Application US/09242737  
Patent No. 6368809  
GENERAL INFORMATION:  
APPLICANT: BERNARDS, REN (

TITLE OF INVENTION: E2F UBIQUITINATION DOMAIN, AND ASSAYS FOR  
INHIBITORS OF E2F UBIQUITINATION  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: NIXON & VANDERHAYE P.C.  
STREET: 1100 NORTH GLEBE ROAD  
CITY: ARLINGTON  
STATE: VIRGINIA  
COUNTRY: U.S.A.  
ZIP: 22201-4714  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/242,737  
FILING DATE: 23-Feb-1999  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/GB97/02293  
FILING DATE: 22-AUG-1997  
APPLICATION NUMBER: GB 9617697.9  
FILING DATE: 23-AUG-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: WILSON, MARY J.  
REGISTRATION NUMBER: 32,955  
REFERENCE/DOCKET NUMBER: 620-63  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 816-4000  
TELEFAX: (703) 816-4100  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 437 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 2:  
US-09-242-737-2

Query Match 88.9%; Score 32; DB 3; Length 437;  
Best Local Similarity 100.0%; Pred. No. 8.3e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGPC 5  
Db 8 AGGPC 12

RESULT 37  
US-09-315-113-1  
Sequence 1, Application US/09315113  
Patent No. 6379927  
GENERAL INFORMATION:  
APPLICANT: Antelman, Douglas  
Gregory, Richard J.  
Wils, Kenneth N.  
TITLE OF INVENTION: Tissue Specific Expression of  
Retinoblastoma Protein  
NUMBER OF SEQUENCES: 46  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP  
STREET: Two Embarcadero Center, 8th Floor  
CITY: San Francisco  
STATE: CA  
COUNTRY: USA  
ZIP: 94111  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:

```

; APPLICATION NUMBER: US/09/315,113
; FILING DATE: 19-MAY-1999
; CLASSIFICATION: <Unknown>
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 08/801,092
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: FITEB, Renee A.
; REGISTRATION NUMBER: 35,136
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-576-0200
; TELEFAX: 703-576-0300
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 437 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-315-113-1

Query Match      88.9%; Score 32; DB 3; Length 437;
Best Local Similarity 100.0%; Pred. No. 8.3e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AGGPG 5
        |||||
Db      8 AGGPG 12

RESULT 38
US-09-710-861-13
; Sequence 13, Application US/09710861
; GENERAL INFORMATION:
; APPLICANT: La Thangue, Nicholas
; TITLE OF INVENTION: TRANSCRIPTION FACTOR DP-3 AND ISOFORMS THEREOF
; FILE REFERENCE: 620-54
; CURRENT APPLICATION NUMBER: US/09/710,861
; PRIOR FILING DATE: 2000-11-13
; PRIOR APPLICATION NUMBER: US/09/189,627
; PRIOR FILING DATE: 1998-11-10
; PRIOR APPLICATION NUMBER: 08/723,415
; PRIOR FILING DATE: 1996-09-30
; PRIOR APPLICATION NUMBER: GB 9610195
; PRIOR FILING DATE: 1996-05-15
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO: 13
; LENGTH: 437
; TYPE: PRT
; ORGANISM: human
US-09-710-861-13

Query Match      88.9%; Score 32; DB 3; Length 437;
Best Local Similarity 100.0%; Pred. No. 8.3e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 AGGPG 5
        |||||
Db      8 AGGPG 12

RESULT 39
US-08-450-360-4
; Sequence 4, Application US/08450360
; Patent No. 5656457
; GENERAL INFORMATION:
; APPLICANT: Parke, Deborah Lynn
; APPLICANT: Coates, Stephen Ralph
```

```

; TITLE OF INVENTION: Herpes Simplex Virus Type 2-Glycoprotein G
; TITLE OF INVENTION: Proteins and Polypeptides
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Leona L. Lauder
; STREET: 6 Mariposa Court
; CITY: Tiburon
; STATE: California
; COUNTRY: USA
; ZIP: 94920
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/450,360
; FILING DATE: 25-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/129,021
; FILING DATE: 29-SEP-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Lauder, Leona L.
; REGISTRATION NUMBER: 30,863
; REFERENCE/DOCKET NUMBER: D-0012.2A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-435-2034
; TELEFAX: 415-435-0727
; GENERAL INFORMATION:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 447 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-450-360-4

Query Match      88.9%; Score 32; DB 1; Length 447;
Best Local Similarity 100.0%; Pred. No. 8.5e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      2 GGPCV 6
        |||||
Db      167 GGPCV 171

RESULT 40
US-08-139-937-14
; Sequence 14, Application US/08139937
; Patent No. 5821070
; GENERAL INFORMATION:
; APPLICANT: LEE, WEN-HWA
; APPLICANT: SHAN, BEI
; TITLE OF INVENTION: CELLULAR GENES ENCODING
; TITLE OF INVENTION: RETINOBLASTOMA-ASSOCIATED PROTEINS
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CAMPBELL AND FLORES
; STREET: 4370 LA JOLLA VILLAGE DRIVE
; CITY: SAN DIEGO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/139,937
; FILING DATE: 20-OCT-1993
; CLASSIFICATION: 435
```



PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/979,156  
FILING DATE: 20-NOV-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: CAMPBELL, CATRYN  
REGISTRATION NUMBER: 31,815  
REFERENCE/DOCKET NUMBER: P-CJ 9370  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619-535-9001  
TELEFAX: 619-535-8949  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 476 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-139-937-14

Query Match 88.9%; Score 32; DB 2; Length 476;  
Best Local Similarity: 100.0%; Pred. No. 9e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGGPC 5  
|||  
Db 47 AGGPC 51

Search completed: December 30, 2004, 15:16:02  
Job time : 7.19102 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: December 30, 2004, 15:14:08 : Search time 0.736782 Seconds  
(without alignments)  
2929.449 Million cell updates/sec

Title: US-10-719-385-26

Perfect score: 36

Sequence: 1 AGGRCV 6

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1599051 seqs, 359727711 residues

Total number of hits satisfying chosen parameters: 1599051

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 150 summaries

Database :

Published Applications AA:\*

1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*

2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep:\*

3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep:\*

4: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:\*

5: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep:\*

6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep:\*

7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep:\*

8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*

9: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*

10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*

11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep:\*

12: /cgn2\_6/ptodata/1/pubpaa/US09C\_NEW\_PUB.pep:\*

13: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep:\*

14: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep:\*

15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep:\*

16: /cgn2\_6/ptodata/1/pubpaa/US10D\_PUBCOMB.pep:\*

17: /cgn2\_6/ptodata/1/pubpaa/US10D\_NEW\_PUB.pep:\*

18: /cgn2\_6/ptodata/1/pubpaa/US11\_NEW\_PUB.pep:\*

19: /cgn2\_6/ptodata/1/pubpaa/US11\_NEW\_PUB.pep:\*

20: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	36	100.0	6 17 US-10-719-385-26	Sequence 26, App1
2	36	100.0	68 17 US-10-425-115-359821	Sequence 359821,
3	36	100.0	339 16 US-10-437-963-157034	Sequence 157034,
4	36	100.0	407 16 US-10-408-765A-817	Sequence 817, App
5	36	100.0	858 14 US-10-177-293-47	Sequence 47, App1
6	36	100.0	1745 17 US-10-719-385-21	Sequence 21, App1
7	36	100.0	1745 17 US-10-370-715B-544	Sequence 544, App
8	36	100.0	3534 17 US-10-719-385-22	Sequence 22, App1
9	35	97.2	206 15 US-10-424-589-230302	Sequence 230302,
10	35	97.2	378 15 US-10-424-589-149625	Sequence 149625,
11	35	97.2	409 15 US-10-424-589-279664	Sequence 279664,
12	35	97.2	1036 9 US-09-771-161A-255	Sequence 255, App
13	35	97.2	1036 9 US-09-771-161A-256	Sequence 256, App

14	35	97.2	1036	14	US-10-318-410-2	Sequence 2, App1
15	35	97.2	1092	17	US-10-737-450-72	Sequence 72, App1
16	33	91.7	82	16	US-10-437-963-201950	Sequence 201950,
17	33	91.7	93	16	US-10-437-963-153250	Sequence 153250,
18	33	91.7	95	16	US-10-437-963-121278	Sequence 121278,
19	33	91.7	98	16	US-10-767-701-51918	Sequence 51918, A
20	33	91.7	149	16	US-10-437-963-170381	Sequence 170381,
21	33	91.7	152	16	US-10-437-963-129834	Sequence 129834,
22	33	91.7	338	17	US-10-425-115-235622	Sequence 235622,
23	33	91.7	351	16	US-10-767-701-40379	Sequence 40379, A
24	33	91.7	355	16	US-10-767-701-41600	Sequence 41600, A
25	33	91.7	371	15	US-10-425-114-52879	Sequence 52879, A
26	33	91.7	782	16	US-10-437-963-180870	Sequence 180870,
27	33	91.7	931	16	US-10-437-963-154670	Sequence 154670,
28	33	91.7	1824	16	US-10-437-963-194642	Sequence 194642,
29	32	88.9	15	14	US-10-354-240-34	Sequence 34, App1
30	32	88.9	15	14	US-10-354-240-35	Sequence 35, App1
31	32	88.9	18	17	US-10-425-115-341764	Sequence 341764,
32	32	88.9	43	16	US-10-437-963-186690	Sequence 186690,
33	32	88.9	46	17	US-10-425-115-249558	Sequence 249558,
34	32	88.9	48	14	US-10-029-386-31343	Sequence 31343, A
35	32	88.9	48	17	US-10-425-115-286513	Sequence 286513,
36	32	88.9	52	15	US-10-424-599-241100	Sequence 241100,
37	32	88.9	54	15	US-10-424-599-181045	Sequence 181045,
38	32	88.9	57	15	US-10-424-599-152102	Sequence 152102,
39	32	88.9	57	17	US-10-425-115-336645	Sequence 336645,
40	32	88.9	62	17	US-10-425-115-259867	Sequence 259867,
41	32	88.9	62	17	US-10-425-115-283116	Sequence 283116,
42	32	88.9	66	17	US-10-425-115-220914	Sequence 220914,
43	32	88.9	68	15	US-10-424-599-158838	Sequence 158838,
44	32	88.9	71	17	US-10-425-115-332685	Sequence 332685,
45	32	88.9	73	15	US-10-424-599-285389	Sequence 285389,
46	32	88.9	76	15	US-10-424-599-239335	Sequence 239335,
47	32	88.9	76	17	US-10-425-115-349179	Sequence 349179,
48	32	88.9	78	17	US-10-425-115-325353	Sequence 325353,
49	32	88.9	81	17	US-10-425-115-320538	Sequence 320538,
50	32	88.9	81	17	US-10-425-115-193365	Sequence 193365,
51	32	88.9	85	17	US-10-425-115-295988	Sequence 295988,
52	32	88.9	86	16	US-10-437-963-134878	Sequence 134878,
53	32	88.9	89	9	US-09-739-907-83	Sequence 83, App1
54	32	88.9	89	11	US-09-938-671-83	Sequence 83, App1
55	32	88.9	91	15	US-10-289-762-71	Sequence 76, App
56	32	88.9	100	17	US-10-425-115-209246	Sequence 209246,
57	32	88.9	102	17	US-10-425-115-303565	Sequence 303565,
58	32	88.9	104	17	US-10-425-115-334205	Sequence 334205,
59	32	88.9	106	15	US-10-424-599-284867	Sequence 284867,
60	32	88.9	114	15	US-10-424-599-146860	Sequence 146860,
61	32	88.9	116	15	US-10-424-599-172385	Sequence 172385,
62	32	88.9	116	17	US-10-425-115-293010	Sequence 293010,
63	32	88.9	118	9	US-09-739-907-182	Sequence 182, App
64	32	88.9	118	11	US-09-938-671-182	Sequence 182, App
65	32	88.9	120	16	US-10-767-701-57956	Sequence 57956, A
66	32	88.9	124	17	US-10-425-115-201607	Sequence 201607,
67	32	88.9	126	15	US-10-424-599-235482	Sequence 235482,
68	32	88.9	132	17	US-10-425-115-185119	Sequence 185119,
69	32	88.9	132	17	US-10-425-115-340087	Sequence 340087,
70	32	88.9	142	17	US-10-425-115-313667	Sequence 313667,
71	32	88.9	143	15	US-10-424-599-252887	Sequence 252887,
72	32	88.9	144	17	US-10-425-115-280549	Sequence 280549,
73	32	88.9	148	17	US-10-425-115-299166	Sequence 299166,
74	32	88.9	152	15	US-10-424-599-251472	Sequence 251472,
75	32	88.9	155	17	US-10-425-115-253866	Sequence 253866,
76	32	88.9	157	16	US-10-437-963-119046	Sequence 119046,
77	32	88.9	164	16	US-10-437-963-119035	Sequence 119035,
78	32	88.9	166	16	US-10-437-963-102623	Sequence 102623,
79	32	88.9	166	16	US-10-437-963-174066	Sequence 174066,
80	32	88.9	174	16	US-10-437-963-119968	Sequence 119968,
81	32	88.9	174	16	US-10-767-701-32294	Sequence 32294, A
82	32	88.9	194	14	US-10-104-047-2315	Sequence 2315, Ap
83	32	88.9	198	9	US-09-995-553A-3	Sequence 5, App1
84	32	88.9	207	15	US-10-425-114-46073	Sequence 46073, A
85	32	88.9	211	16	US-10-767-701-35984	Sequence 35984, A
86	32	88.9	216	15	US-10-108-260A-4474	Sequence 4474, Ap

87	32	88.9	229	14	US-10-156-761-12171	Sequence 12171, A
88	32	88.9	247	15	US-10-424-599-272621	Sequence 272621, A
89	32	88.9	248	14	US-10-121-988-142	Sequence 142, App
90	32	88.9	248	14	US-10-200-562-142	Sequence 142, App
91	32	88.9	248	14	US-10-237-551-142	Sequence 142, App
92	32	88.9	249	17	US-10-739-930-7995	Sequence 7995, App
93	32	88.9	253	15	US-10-108-260A-3556	Sequence 3556, App
94	32	88.9	266	14	US-10-104-047-3114	Sequence 3114, App
95	32	88.9	269	16	US-10-437-963-169997	Sequence 169997, App
96	32	88.9	270	15	US-10-339-148A-7	Sequence 7, Appli
97	32	88.9	285	17	US-10-493-462-31	Sequence 31, Appli
98	32	88.9	294	15	US-10-425-114-48469	Sequence 48469, A
99	32	88.9	296	17	US-10-763-362-6	Sequence 6, Appli
100	32	88.9	298	16	US-10-437-963-140406	Sequence 140406, A
101	32	88.9	320	17	US-10-425-115-358651	Sequence 358651, A
102	32	88.9	323	16	US-10-437-963-185078	Sequence 185078, A
103	32	88.9	332	15	US-10-425-114-66862	Sequence 66862, A
104	32	88.9	338	16	US-10-437-963-186313	Sequence 186313, A
105	32	88.9	351	14	US-10-104-047-2926	Sequence 2926, App
106	32	88.9	352	14	US-10-156-761-9067	Sequence 9067, App
107	32	88.9	362	15	US-10-425-114-62852	Sequence 62852, A
108	32	88.9	374	10	US-09-847-208-68	Sequence 68, Appli
109	32	88.9	393	17	US-10-739-930-10911	Sequence 10911, A
110	32	88.9	407	14	US-10-157-031-281	Sequence 281, App
111	32	88.9	424	13	US-10-029-180-104	Sequence 104, App
112	32	88.9	424	13	US-10-149-310-302	Sequence 302, App
113	32	88.9	430	16	US-10-767-701-45374	Sequence 45374, A
114	32	88.9	436	17	US-10-425-115-330241	Sequence 330241, A
115	32	88.9	446	15	US-10-276-774-1370	Sequence 1370, App
116	32	88.9	462	15	US-10-282-122A-76568	Sequence 76568, A
117	32	88.9	531	14	US-10-369-493-2589	Sequence 2589, App
118	32	88.9	535	14	US-10-369-493-22630	Sequence 22630, A
119	32	88.9	581	14	US-10-094-466-40	Sequence 40, Appli
120	32	88.9	687	14	US-10-369-493-5922	Sequence 5922, App
121	32	88.9	699	14	US-10-121-988-143	Sequence 143, App
122	32	88.9	699	14	US-10-200-562-143	Sequence 143, App
123	32	88.9	699	14	US-10-237-551-143	Sequence 143, App
124	32	88.9	699	14	US-10-237-551-254	Sequence 254, App
125	32	88.9	719	14	US-10-156-761-9077	Sequence 9077, App
126	32	88.9	735	13	US-10-068-674-4	Sequence 4, Appli
127	32	88.9	773	16	US-10-437-963-121782	Sequence 121782, A
128	32	88.9	789	16	US-10-437-963-145233	Sequence 145233, A
129	32	88.9	799	15	US-10-236-392-58	Sequence 58, Appli
130	32	88.9	804	14	US-10-369-493-10624	Sequence 10624, A
131	32	88.9	878	15	US-10-398-197-1	Sequence 1, Appli
132	32	88.9	969	15	US-10-094-886-140	Sequence 140, App
133	32	88.9	1036	9	US-09-995-593A-6	Sequence 6, Appli
134	32	88.9	1067	9	US-10-650-650-18	Sequence 18, Appli
135	32	88.9	1187	9	US-09-995-593A-7	Sequence 7, Appli
136	32	88.9	1193	17	US-10-877-563-8	Sequence 8, Appli
137	32	88.9	1208	14	US-10-213-329-1	Sequence 1, Appli
138	32	88.9	1218	9	US-09-855-722-7	Sequence 7, Appli
139	32	88.9	1218	9	US-09-995-593A-11	Sequence 11, Appli
140	32	88.9	1218	9	US-09-944-849-3	Sequence 3, Appli
141	32	88.9	1218	14	US-10-219-248-7	Sequence 7, Appli
142	32	88.9	1218	14	US-10-219-247-7	Sequence 7, Appli
143	32	88.9	1218	14	US-10-281-478-7	Sequence 7, Appli
144	32	88.9	1218	15	US-10-235-392-56	Sequence 56, Appli
145	32	88.9	1218	16	US-10-648-593-155	Sequence 155, App
146	32	88.9	1218	17	US-10-763-362-43	Sequence 43, Appli
147	32	88.9	1218	17	US-10-877-563-4	Sequence 4, Appli
148	32	88.9	1218	17	US-10-650-650-1	Sequence 1, Appli
149	32	88.9	1404	9	US-09-844-849-8	Sequence 8, Appli
150	32	88.9	4840	14	US-10-156-761-10435	Sequence 10435, A

## ALIGNMENTS

RESULT 1  
US-10-719-385-26  
Sequence 26, Application US/10719385  
Publication No. US20040209284A1

GENERAL INFORMATION:  
APPLICANT: O'Toole et al.  
TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis  
FILE REFERENCE: 22058-582  
CURRENT FILING DATE: 2003-11-21  
PRIOR APPLICATION NUMBER: PCT/US03/37339  
PRIOR FILING DATE: 2003-11-21  
PRIOR APPLICATION NUMBER: 60/428,094  
NUMBER OF SEQ ID NOS: 26  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 26  
LENGTH: 6  
TYPE: PRT  
ORGANISM: Homo sapiens

Query Match 100.0%; Score 36; DB 17; Length 6;  
Best Local Similarity 100.0%; Pred. No. 1.5e+06;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGPCV 6  
Db 1 AGPCV 6

RESULT 2  
US-10-425-115-359921  
Sequence 359921, Application US/10425115  
Publication No. US20040214272A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa, Thomas J.  
APPLICANT: Kovacic, David K.  
APPLICANT: Zhou, Yihua  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(53222)B  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 369326  
SEQ ID NO 359921  
LENGTH: 68  
TYPE: PRT  
ORGANISM: Zea mays  
FEATURE:  
OTHER INFORMATION: Clone ID: MFT4577\_9141C.1.pap  
US-10-425-115-359921

Query Match 100.0%; Score 36; DB 17; Length 68;  
Best Local Similarity 100.0%; Pred. No. 87;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGPCV 6  
Db 38 AGPCV 43

RESULT 3  
US-10-437-963-157034  
Sequence 157034, Application US/10437963  
Publication No. US20040123343A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa, Thomas J.  
APPLICANT: Kovacic, David K.  
APPLICANT: Zhou, Yihua  
APPLICANT: Cao, Yongwei  
APPLICANT: Wu, Wei  
APPLICANT: Boukharov, Andrey A.  
APPLICANT: Barbazuk, Brad  
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With

```

; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 157034
; LENGTH: 339
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: clone ID: PAT_MKT4530_56644C.1.Dep
US-10-437-963-157034

Query Match
Best Local Similarity 100.0%; Score 36; DB 16; Length 339;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 AGPCV 6
Db 6 AGPCV 11

RESULT 4
US-10-408-765A-817
; Sequence 817, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary W.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; CURRENT FILING DATE: 2003-04-04
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 817
; LENGTH: 407
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-817

Query Match
Best Local Similarity 100.0%; Score 36; DB 16; Length 407;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 AGPCV 6
Db 280 AGPCV 285

RESULT 5
US-10-177-293-47
; Sequence 47, Application US/10177293
; Publication No. US20030124128A1
; GENERAL INFORMATION:
; APPLICANT: Little, James
; APPLICANT: Gillet, Karen
; APPLICANT: Zhao, Xumei
; APPLICANT: Gannavairpu, Manjula
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Mertens, Maureen
; APPLICANT: Myer, Vic
; APPLICANT: Wang, Youzhen
; APPLICANT: Xu, Yongyao
; APPLICANT: Hoersch, Sebastian
; APPLICANT: Monahan, John
; APPLICANT: Meyers, Rachel E.

; APPLICANT: Baat Jr., Robert C.
; APPLICANT: Hortobagyi, Gabriel N.
; APPLICANT: Puzstai, Lajos
; APPLICANT: Meric, Funda
; APPLICANT: Sahin, Aysegul
; APPLICANT: Mills, Gordon B.
; TITLE OF INVENTION: COMPOSITIONS, KITS, AND METHODS FOR IDENTIFICATION, ASSESSMENT,
; PREVENTION, AND THERAPY OF BREAST CANCER
; FILE REFERENCE: MRI-038
; CURRENT APPLICATION NUMBER: US/10/177,293
; CURRENT FILING DATE: 2002-06-21
; PRIOR APPLICATION NUMBER: US 60/299,887
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 60/301,572
; PRIOR FILING DATE: 2001-06-27
; PRIOR APPLICATION NUMBER: US 60/306,501
; PRIOR FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: US 60/325,002
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 60/362,585
; PRIOR FILING DATE: 2002-03-05
; PRIOR APPLICATION NUMBER: US 60/xxx,xxx
; PRIOR FILING DATE: 2002-05-14
; NUMBER OF SEQ ID NOS: 506
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 47
; LENGTH: 858
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-177-293-47

Query Match
Best Local Similarity 100.0%; Score 36; DB 14; Length 858;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 AGPCV 6
Db 731 AGPCV 736

RESULT 6
US-10-719-385-21
; Sequence 21, Application US/10719385
; Publication No. US20040209284A1
; GENERAL INFORMATION:
; APPLICANT: O'Toole et al.
; TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis
; FILE REFERENCE: 22058-582
; CURRENT APPLICATION NUMBER: US/10/719,385
; CURRENT FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: PCT/US03/37339
; PRIOR FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: 60/428,094
; PRIOR FILING DATE: 2002-11-21
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 21
; LENGTH: 1745
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-385-21

Query Match
Best Local Similarity 100.0%; Score 36; DB 17; Length 1745;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Cy 1 AGPCV 6
Db 1 AGPCV 6

RESULT 7
US-10-370-715B-544
```

Sequence 544, Application US/10370715B  
Publication No. US20040258678A1  
GENERAL INFORMATION:  
Patin Docket Preview  
APPLICANT: BODARY, SARAH C.  
APPLICANT: CLARK, HILARY  
APPLICANT: BRISDELL, HUNTE  
APPLICANT: JACKMAN, JANET  
APPLICANT: SCHOENFELD, JILL R.  
APPLICANT: WILLIAMS, P. MICKEY  
APPLICANT: WOOD, WILLIAM I.  
APPLICANT: WU, THOMAS D.  
TITLE OF INVENTION: Compositions and Methods for the Treatment of Immune  
TITLE OF INVENTION: Related Diseases  
FILE REFERENCE: P1948P1-US  
CURRENT APPLICATION NUMBER: US/10/370,715B  
CURRENT FILING DATE: 2003-02-21  
NUMBER OF SEQ ID NOS: 742  
SEQ ID NO 544  
LENGTH: 1745  
TYPE: PRT  
ORGANISM: Homo sapien  
US-10-370-715B-544

Query Match 100.0%; Score 36; DB 17; Length 1745;  
Best Local Similarity 100.0%; Pred. No. 1.4e+03;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGPCV 6  
111111  
Db 1 AGPCV 6

RESULT 8  
US-10-719-385-22  
Sequence 22, Application US/10719385  
Publication No. US20040209284A1  
GENERAL INFORMATION:  
APPLICANT: O'Toole et al.  
TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis  
FILE REFERENCE: 22058-582  
CURRENT APPLICATION NUMBER: US/10/719,385  
CURRENT FILING DATE: 2003-11-21  
PRIOR APPLICATION NUMBER: PCT/US03/37339  
PRIOR FILING DATE: 2003-11-21  
PRIOR APPLICATION NUMBER: 60/428,094  
PRIOR FILING DATE: 2002-11-21  
NUMBER OF SEQ ID NOS: 26  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 22  
LENGTH: 3534  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-719-385-22

Query Match 100.0%; Score 36; DB 17; Length 3534;  
Best Local Similarity 100.0%; Pred. No. 2.6e+03;  
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGPCV 6  
111111  
Db 19 AGPCV 24

RESULT 9  
US-10-424-599-230302  
Sequence 230302, Application US/10424599  
Publication No. US20040031072A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa, Thomas J  
APPLICANT: Kovalic, David K  
APPLICANT: Zhou Yihua  
APPLICANT: Cao Yongwei

TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21(53223)B  
CURRENT APPLICATION NUMBER: US/10/424,599  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 285684  
SEQ ID NO 230302  
LENGTH: 206  
TYPE: PRT  
ORGANISM: Glycine max  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_49987C.1.pep  
US-10-424-599-230302

Query Match 97.2%; Score 35; DB 15; Length 206;  
Best Local Similarity 83.3%; Pred. No. 3.3e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGPCV 6  
111111  
Db 159 AGPCV 164

RESULT 10  
US-10-424-599-149825  
Sequence 149825, Application US/10424599  
Publication No. US20040031072A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa, Thomas J  
APPLICANT: Kovalic, David K  
APPLICANT: Zhou Yihua  
APPLICANT: Cao Yongwei  
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21(53223)B  
CURRENT APPLICATION NUMBER: US/10/424,599  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 285684  
SEQ ID NO 149825  
LENGTH: 378  
TYPE: PRT  
ORGANISM: Glycine max  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (1) (378)  
OTHER INFORMATION: unsure at all Xaa locations  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_106313C.1.pep  
US-10-424-599-149825

Query Match 97.2%; Score 35; DB 15; Length 378;  
Best Local Similarity 83.3%; Pred. No. 5.6e+02;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGPCV 6  
111111  
Db 131 AGPCV 136

RESULT 11  
US-10-424-599-279664  
Sequence 279664, Application US/10424599  
Publication No. US20040031072A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa, Thomas J  
APPLICANT: Kovalic, David K  
APPLICANT: Zhou Yihua  
APPLICANT: Cao Yongwei  
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21(53223)B  
CURRENT APPLICATION NUMBER: US/10/424,599  
CURRENT FILING DATE: 2003-04-28

NUMBER OF SEQ ID NOS: 285684  
SEQ ID NO 279664  
LENGTH: 409  
TYPE: PRT  
ORGANISM: Glycine max  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (1)..(409)  
OTHER INFORMATION: unsure at all Xaa locations  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MKT3847\_94559C.1.Dep  
US-10-424-599-279664

Query Match 97.2% Score 35; DB 15; Length 409;  
Best Local Similarity 83.3% Pred. No. 1.3e+03;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGPCV 6  
Db 161 AGPCV 166

RESULT 12  
US-09-771-161A-255  
Sequence 255, Application US/09771161A  
Patent No. US20020110811A1  
GENERAL INFORMATION:  
APPLICANT: LEVINE, et al.  
TITLE OF INVENTION: VARIANTS OF PROTEIN KINASES  
FILE REFERENCE: 802620-2005.1  
CURRENT FILING DATE: 2001-01-26  
CURRENT APPLICATION NUMBER: US/09/771,161A  
PRIOR FILING DATE: 2000-11-28  
PRIOR APPLICATION NUMBER: 09/724,676  
PRIOR FILING DATE: 2000-06-15  
PRIOR APPLICATION NUMBER: 135619  
PRIOR FILING DATE: 2000-04-12  
NUMBER OF SEQ ID NOS: 273  
SOFTWARE: Patentin version 3.0  
SEQ ID NO 255  
LENGTH: 1036  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-771-161A-255

Query Match 97.2% Score 35; DB 9; Length 1036;  
Best Local Similarity 83.3% Pred. No. 1.3e+03;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGPCV 6  
Db 837 AGPCV 842

RESULT 13  
US-09-771-161A-256  
Sequence 256, Application US/09771161A  
Patent No. US20020110811A1  
GENERAL INFORMATION:  
APPLICANT: LEVINE, et al.  
TITLE OF INVENTION: VARIANTS OF PROTEIN KINASES  
FILE REFERENCE: 802620-2005.1  
CURRENT FILING DATE: 2001-01-26  
CURRENT APPLICATION NUMBER: US/09/771,161A  
PRIOR FILING DATE: 2000-11-28  
PRIOR APPLICATION NUMBER: 09/724,676  
PRIOR FILING DATE: 2000-06-15  
PRIOR APPLICATION NUMBER: 135619  
PRIOR FILING DATE: 2000-04-12  
NUMBER OF SEQ ID NOS: 273  
SOFTWARE: Patentin version 3.0

SEQ ID NO 256  
LENGTH: 1036  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-771-161A-256

Query Match 97.2% Score 35; DB 9; Length 1036;  
Best Local Similarity 83.3% Pred. No. 1.3e+03;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGPCV 6  
Db 837 AGPCV 842

RESULT 14  
US-10-318-410-2  
Sequence 2, Application US/10318410  
Publication No. US20030175763A1  
GENERAL INFORMATION:  
APPLICANT: Yan Y. DEGENHARDT  
APPLICANT: Scott POWERS  
TITLE OF INVENTION: IDENTIFICATION OF AN AMPLIFIED GENE AND TARGET FOR DRUG  
FILE REFERENCE: 38002-0040  
CURRENT FILING DATE: 2002-12-13  
PRIOR APPLICATION NUMBER: US/10/318,410  
PRIOR FILING DATE: 2001-12-20  
NUMBER OF SEQ ID NOS: 16  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 2  
LENGTH: 1036  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-318-410-2

Query Match 97.2% Score 35; DB 14; Length 1036;  
Best Local Similarity 83.3% Pred. No. 1.3e+03;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGPCV 6  
Db 837 AGPCV 842

RESULT 15  
US-10-737-450-72  
Sequence 72, Application US/10737450  
Publication No. US20040235071A1  
GENERAL INFORMATION:  
APPLICANT: Millennium Pharmaceuticals, Inc  
APPLICANT: Lightcap, Eric S.  
APPLICANT: Ecsedy, Jeffrey A.  
APPLICANT: Hunter, John Joseph  
APPLICANT: Macbeth, Kyle J  
APPLICANT: Tighe Nestor, Michelle  
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR TREATING  
TITLE OF INVENTION: CANCER USING 15986, 2188, 20743, 9148, 9151, 9791, 44252,  
TITLE OF INVENTION: 14184, 42461, 8204, 7970, 25552, 21657, 26492, 2411, 15088,  
TITLE OF INVENTION: 1908, 28899, 63380, 33935, 10480, 12686, 25501, 17694,  
TITLE OF INVENTION: 15701, 53062, 49908, 21612, 38949, 6216, 46863, 9235, 2201,  
TITLE OF INVENTION: 6985, 9883, 12238, 18057, 21617, 39228, 49928, 54476, 62113,  
TITLE OF INVENTION: 64316, 12264, 32362, 58198, 2887, 3205, 8557, 9600, 9693,  
TITLE OF INVENTION: 44867, 53058, 55556, 57658, 2208, 10252, 10302, 14218,  
TITLE OF INVENTION: 33877, 10317, 10485, 25964, 14815, 1363, 1397, 14827, 21708,  
FILE REFERENCE: MP102-207BIRNOMIM  
CURRENT FILING DATE: 2003-12-16  
CURRENT APPLICATION NUMBER: US/10/737,450  
PRIOR FILING DATE: 2002-12-20  
PRIOR APPLICATION NUMBER: US 60/435,108  
PRIOR FILING DATE: 2002-12-20  
PRIOR APPLICATION NUMBER: US 60/436,443

```

; PRIOR FILING DATE: 2002-12-23
; PRIOR APPLICATION NUMBER: US 60/438,498
; PRIOR FILING DATE: 2003-01-07
; PRIOR APPLICATION NUMBER: US 60/444,370
; PRIOR FILING DATE: 2003-01-31
; PRIOR APPLICATION NUMBER: US 60/446,031
; PRIOR FILING DATE: 2003-02-06
; PRIOR APPLICATION NUMBER: US 60/453,635
; PRIOR FILING DATE: 2003-03-11
; PRIOR APPLICATION NUMBER: US 60/457,199
; PRIOR FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: US 60/462,458
; PRIOR FILING DATE: 2003-04-10
; PRIOR APPLICATION NUMBER: US 60/466,732
; PRIOR FILING DATE: 2003-04-30
; PRIOR APPLICATION NUMBER: US 60/469,184
; PRIOR FILING DATE: 2003-05-08
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 144
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 72
; LENGTH: 1092
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-10-737-450-72

Query Match      97.2%; Score 35; DB 17; Length 1092;
Best Local Similarity 83.3%; Pred. No. 1.4e+03;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy      1 AGPCV 6
        |||||:
Db      893 AGPCL 898

RESULT 16
US-10-437-963-201950
; Sequence 201950, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 201950
; LENGTH: 82
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(82)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_97276C.1.pep
US-10-437-963-201950

Query Match      91.7%; Score 33; DB 16; Length 82;
Best Local Similarity 83.3%; Pred. No. 3.2e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy      1 AGPCV 6
        |||||:
Db      15 AGPCL 20

US-10-437-963-153250
; Sequence 153250, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 153250
; LENGTH: 93
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_53220C.1.pep
US-10-437-963-153250

Query Match      91.7%; Score 33; DB 16; Length 93;
Best Local Similarity 83.3%; Pred. No. 3.6e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy      1 AGPCV 6
        |||||:
Db      80 AGPCL 85

RESULT 18
US-10-437-963-121278
; Sequence 121278, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 121278
; LENGTH: 95
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_24319C.1.pep
US-10-437-963-121278

Query Match      91.7%; Score 33; DB 16; Length 95;
Best Local Similarity 83.3%; Pred. No. 3.6e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy      1 AGPCV 6
        |||||:
Db      70 AGPCL 75
```



```
RESULT 19
US-10-767-701-51998
; Sequence 51998, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53535)B
; CURRENT APPLICATION NUMBER: US/10/767,701
; CURRENT FILING DATE: 2004-01-29
; NUMBER OF SEQ ID NOS: 63128
; SEQ ID NO 51998
; LENGTH: 98
; TYPE: PRT
; ORGANISM: Sorghum bicolor
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB5121-002-A1-PF1-E10.pep
US-10-767-701-51998

Query Match          91.7%; Score 33; DB 16; Length 98;
Best Local Similarity 83.3%; Pred. No. 3.7e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY      1 AGPCV 6
      |||||:
Db      32 AGPCL 37

RESULT 20
US-10-437-963-170381
; Sequence 170381, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 170381
; LENGTH: 149
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_68713C.1.pcp
US-10-437-963-170381

Query Match          91.7%; Score 33; DB 16; Length 149;
Best Local Similarity 83.3%; Pred. No. 5.4e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY      1 AGPCV 6
      |||||:
Db      80 AGPCL 85

RESULT 21
US-10-437-963-129834
; Sequence 129834, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
```

```
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 129834
; LENGTH: 152
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(152)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_32052C.1.pcp
US-10-437-963-129834

Query Match          91.7%; Score 33; DB 16; Length 152;
Best Local Similarity 83.3%; Pred. No. 5.5e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY      1 AGPCV 6
      |||||:
Db      84 AGPCL 89

RESULT 22
US-10-425-115-235622
; Sequence 235622, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 235622
; LENGTH: 338
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_146471C.1.pcp
US-10-425-115-235622

Query Match          91.7%; Score 33; DB 17; Length 338;
Best Local Similarity 83.3%; Pred. No. 1.1e+03;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY      1 AGPCV 6
      |||||:
Db      9 AGPCL 14

RESULT 23
US-10-767-701-40379
; Sequence 40379, Application US/10767701
; Publication No. US20040172684A1
; GENERAL INFORMATION:
; APPLICANT: Kovalic, David K.
```

APPLICANT: Zhou, Yihua  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(53535)B  
CURRENT FILING DATE: 2004-01-29  
NUMBER OF SEQ ID NOS: 63128  
SEQ ID NO 40379  
LENGTH: 351  
TYPE: PRT  
ORGANISM: Sorghum bicolor  
FEATURE:  
OTHER INFORMATION: Clone ID: SORBI-28WAY03-C58371\_1.pep  
US-10-767-701-40379

Query Match 91.7%; Score 33; DB 16; Length 351;  
Best Local Similarity 83.3%; Pred. No. 1.1e+03;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 AGPCV 6  
Db 8 AGPCV 13

RESULT 24  
US-10-767-701-41600  
Sequence 41600, Application US/10767701  
Publication No. US20040172684A1  
GENERAL INFORMATION:  
APPLICANT: Kovalic, David K.  
APPLICANT: Zhou, Yihua  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(53535)B  
CURRENT FILING DATE: 2004-01-29  
NUMBER OF SEQ ID NOS: 63128  
SEQ ID NO 41600  
LENGTH: 355  
TYPE: PRT  
ORGANISM: Sorghum bicolor  
FEATURE:  
OTHER INFORMATION: Clone ID: SORBI-28WAY03-C50197\_1.pep  
US-10-767-701-41600

Query Match 91.7%; Score 33; DB 16; Length 355;  
Best Local Similarity 83.3%; Pred. No. 1.1e+03;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 AGPCV 6  
Db 9 AGPCV 14

RESULT 25  
US-10-425-114-52879  
Sequence 52879, Application US/10425114  
Publication No. US20040034888A1  
GENERAL INFORMATION:  
APPLICANT: Liu, Jingdong  
APPLICANT: Zhou, Yihua  
APPLICANT: Kovalic, David K.  
APPLICANT: Screen, Steven E.  
APPLICANT: Tabaska, Jack E.  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(5313)B  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 73128

SEQ ID NO 52879  
LENGTH: 371  
TYPE: PRT  
ORGANISM: Zea mays  
FEATURE:  
OTHER INFORMATION: Clone ID: LIB3060-051-G8\_F11.pep  
US-10-425-114-52879

Query Match 91.7%; Score 33; DB 15; Length 371;  
Best Local Similarity 83.3%; Pred. No. 1.2e+03;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 AGPCV 6  
Db 42 AGPCV 47

RESULT 26  
US-10-437-963-180870  
Sequence 180870, Application US/10437963  
Publication No. US20040123343A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa, Thomas J.  
APPLICANT: Kovalic, David K.  
APPLICANT: Zhou, Yihua  
APPLICANT: Cao, Yongwei  
APPLICANT: Wu, Wei  
APPLICANT: Boukharov, Andrey A.  
APPLICANT: Barbazuk, Brad  
APPLICANT: Li, Ping  
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(53221)B  
CURRENT FILING DATE: 2003-05-14  
NUMBER OF SEQ ID NOS: 204966  
SEQ ID NO 180870  
LENGTH: 782  
TYPE: PRT  
ORGANISM: Oryza sativa  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT4530\_7819C.1.pep  
US-10-437-963-180870

Query Match 91.7%; Score 33; DB 16; Length 782;  
Best Local Similarity 83.3%; Pred. No. 2.2e+03;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 AGPCV 6  
Db 6 AGPCV 11

RESULT 27  
US-10-437-963-156470  
Sequence 156470, Application US/10437963  
Publication No. US20040123343A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa, Thomas J.  
APPLICANT: Kovalic, David K.  
APPLICANT: Zhou, Yihua  
APPLICANT: Cao, Yongwei  
APPLICANT: Wu, Wei  
APPLICANT: Boukharov, Andrey A.  
APPLICANT: Barbazuk, Brad  
APPLICANT: Li, Ping  
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(53221)B  
CURRENT FILING DATE: 2003-05-14  
NUMBER OF SEQ ID NOS: 204966  
SEQ ID NO 156470

LENGTH: 931  
TYPE: PRT  
ORGANISM: Oryza sativa  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (1)..(931)  
OTHER INFORMATION: unsure at all Xaa locations  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT4530\_56134C.1.pcp  
US-10-437-963-156470

Query Match 91.7%; Score 33; DB 16; Length 931;  
Best Local Similarity 83.3%; Pred. No. 2.6e+03;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGPCV 6  
DB 307 AGPCV 312

RESULT 28  
US-10-437-963-194642  
Sequence 194642, Application US/10437963  
Publication No. US20040123343A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa, Thomas J.  
APPLICANT: Zhou, Yihua  
APPLICANT: Cao, Yongwei  
APPLICANT: Wu, Wei  
APPLICANT: Boukharov, Andrey A.  
APPLICANT: Barbazuk, Brad  
APPLICANT: Li, Ping  
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated with  
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21(53221)B  
CURRENT APPLICATION NUMBER: US/10/437,963  
CURRENT FILING DATE: 2003-05-14  
NUMBER OF SEQ ID NOS: 204966  
SEQ ID NO 194642  
LENGTH: 1824  
TYPE: PRT  
ORGANISM: Oryza sativa  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (1)..(1824)  
OTHER INFORMATION: unsure at all Xaa locations  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT4530\_90668C.1.pcp  
US-10-437-963-194642

Query Match 91.7%; Score 33; DB 16; Length 1824;  
Best Local Similarity 83.3%; Pred. No. 4.6e+03;  
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 AGPCV 6  
DB 1775 SGGPCV 1780

RESULT 29  
US-10-354-240-34  
Sequence 34, Application US/10354240  
Publication No. US20030185847A1  
GENERAL INFORMATION:  
APPLICANT: Sone, Toshio  
APPLICANT: Kume, Akinori  
APPLICANT: Dairiki, Kazuo  
APPLICANT: Iwama, Akiyo  
APPLICANT: Kino, Kohsuke  
TITLE OF INVENTION: Peptide-Based Immunotherapeutic Agent for Treating Allergic Disease  
FILE REFERENCE: SPO-103D1  
CURRENT APPLICATION NUMBER: US/10/354,240

CURRENT FILING DATE: 2003-01-29  
PRIOR APPLICATION NUMBER: PCT/JP97/00740  
PRIOR FILING DATE: 1997-03-10  
PRIOR APPLICATION NUMBER: US 09/142,524  
PRIOR FILING DATE: 1998-09-09  
NUMBER OF SEQ ID NOS: 174  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 34  
LENGTH: 15  
TYPE: PRT  
ORGANISM: Cryptomeria japonica  
FEATURE:  
NAME/KEY: MISC FEATURE  
LOCATION: (1)..(15)  
OTHER INFORMATION: Cryj1 peptide, Figure 1, Row 120  
US-10-354-240-34

Query Match 88.9%; Score 32; DB 14; Length 15;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GGPCV 6  
DB 9 GGPCV 13

RESULT 30  
US-10-354-240-35  
Sequence 35, Application US/10354240  
Publication No. US20030185847A1  
GENERAL INFORMATION:  
APPLICANT: Sone, Toshio  
APPLICANT: Dairiki, Kazuo  
APPLICANT: Iwama, Akiyo  
APPLICANT: Kino, Kohsuke  
TITLE OF INVENTION: Peptide-Based Immunotherapeutic Agent for Treating Allergic Disease  
FILE REFERENCE: SPO-103D1  
CURRENT APPLICATION NUMBER: US/10/354,240  
CURRENT FILING DATE: 2003-01-29  
PRIOR APPLICATION NUMBER: PCT/JP97/00740  
PRIOR FILING DATE: 1997-03-10  
PRIOR APPLICATION NUMBER: US 09/142,524  
PRIOR FILING DATE: 1998-09-09  
NUMBER OF SEQ ID NOS: 174  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 35  
LENGTH: 15  
TYPE: PRT  
ORGANISM: Cryptomeria japonica  
FEATURE:  
NAME/KEY: MISC FEATURE  
LOCATION: (1)..(15)  
OTHER INFORMATION: Cryj1 peptide, Figure 1, Row 21  
US-10-354-240-35

Query Match 88.9%; Score 32; DB 14; Length 15;  
Best Local Similarity 100.0%; Pred. No. 1.1e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 GGPCV 6  
DB 4 GGPCV 8

RESULT 31  
US-10-425-115-341764  
Sequence 341764, Application US/10425115  
Publication No. US20040214272A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa, Thomas J.  
APPLICANT: Kovalic, David K.  
APPLICANT: Zhou, Yihua

APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(53222)B  
CURRENT APPLICATION NUMBER: US/10/425,115  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 369326  
SEQ ID NO 341764  
LENGTH: 38  
TYPE: PRT  
ORGANISM: Zea mays  
FEATURE:  
OTHER INFORMATION: Clone ID: MRT4577\_74857C.1.pep  
US-10-425-115-341764

Query Match  
Best Local Similarity 88.9%; Score 32; DB 17; Length 38;  
Best Local Similarity 100.0%; Pred. No. 2.4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGPCV 6  
Db 8 GGPCV 12

RESULT 32  
US-10-437-963-186690  
Sequence 186690, Application US/10437963  
Publication No. US20040123343A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa, Thomas J.  
APPLICANT: Kovalic, David K.  
APPLICANT: Zhou, Yihua  
APPLICANT: Cao, Yongwei  
APPLICANT: Mu, Wei  
APPLICANT: Bouharov, Andrey A.  
APPLICANT: Barbazuk, Brad  
APPLICANT: Li, Ping  
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(53221)B  
CURRENT APPLICATION NUMBER: US/10/437,963  
CURRENT FILING DATE: 2003-05-14  
NUMBER OF SEQ ID NOS: 204966  
SEQ ID NO 186690  
LENGTH: 43  
TYPE: PRT  
ORGANISM: Oryza sativa  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT4530\_83466C.1.pep  
US-10-437-963-186690

Query Match  
Best Local Similarity 88.9%; Score 32; DB 16; Length 43;  
Best Local Similarity 100.0%; Pred. No. 2.7e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGPC 5  
Db 25 AGPC 29

RESULT 33  
US-10-425-115-249558  
Sequence 249558, Application US/10425115  
Publication No. US20040214272A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa, Thomas J.  
APPLICANT: Kovalic, David K.  
APPLICANT: Zhou, Yihua  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(53222)B  
CURRENT APPLICATION NUMBER: US/10/425,115

CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 369326  
SEQ ID NO 249558  
LENGTH: 46  
TYPE: PRT  
ORGANISM: Zea mays  
FEATURE:  
OTHER INFORMATION: Clone ID: MRT4577\_159181C.1.pep  
US-10-425-115-249558

Query Match  
Best Local Similarity 88.9%; Score 32; DB 17; Length 46;  
Best Local Similarity 100.0%; Pred. No. 2.9e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGPCV 6  
Db 39 GGPCV 43

RESULT 34  
US-10-029-386-31343  
Sequence 31343, Application US/10029386  
Publication No. US20030194704A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharon G.  
APPLICANT: Hanzel, David R.  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G  
FILE REFERENCE: AEOMICA-X-2  
CURRENT APPLICATION NUMBER: US/10/029,386  
CURRENT FILING DATE: 2001-12-20  
NUMBER OF SEQ ID NOS: 34288  
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 31343  
LENGTH: 48  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AC012384.16  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.2  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.6  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.4  
OTHER INFORMATION: SWISSPROT HIT: P21851, VALUE 1.70e+00  
US-10-029-386-31343

Query Match  
Best Local Similarity 88.9%; Score 32; DB 14; Length 48;  
Best Local Similarity 100.0%; Pred. No. 3e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGPCV 6  
Db 39 GGPCV 43

RESULT 35  
US-10-425-115-286513  
Sequence 286513, Application US/10425115  
Publication No. US20040214272A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa, Thomas J.  
APPLICANT: Kovalic, David K.  
APPLICANT: Zhou, Yihua  
APPLICANT: Cao, Yongwei  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(53222)B  
CURRENT APPLICATION NUMBER: US/10/425,115  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 369326  
SEQ ID NO 286513  
LENGTH: 48  
TYPE: PRT

ORGANISM: Zea mays  
FEATURE:  
OTHER INFORMATION: Clone ID: MRT4577\_24398C.1.pep  
US-10-425-115-286513

Query Match 88.9%; Score 32; DB 17; Length 48;  
Best Local Similarity 66.7%; Pred. No. 3e+02;  
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGPCV 6  
DB 20 SGPCV 25

RESULT 36  
US-10-424-599-241100  
Sequence 241100, Application US/10424599  
Publication No. US20040031072A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa Thomas J  
APPLICANT: Kovalic David K  
APPLICANT: Zhou Yihua  
APPLICANT: Cao Yongwei  
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21(53223)B  
CURRENT APPLICATION NUMBER: US/10/424,599  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 285684  
SEQ ID NO 241100  
LENGTH: 52  
TYPE: PRT  
ORGANISM: Glycine max  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_59741C.1.pep  
US-10-424-599-241100

Query Match 88.9%; Score 32; DB 15; Length 52;  
Best Local Similarity 100.0%; Pred. No. 3.2e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGPCV 5  
DB 38 AGPCV 42

RESULT 37  
US-10-424-599-181045  
Sequence 181045, Application US/10424599  
Publication No. US20040031072A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa Thomas J  
APPLICANT: Kovalic David K  
APPLICANT: Zhou Yihua  
APPLICANT: Cao Yongwei  
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21(53223)B  
CURRENT APPLICATION NUMBER: US/10/424,599  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 285684  
SEQ ID NO 181045  
LENGTH: 54  
TYPE: PRT  
ORGANISM: Glycine max  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (1) (54)  
OTHER INFORMATION: unsure at all Xaa locations  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_13449C.1.pep  
US-10-424-599-181045

Query Match 88.9%; Score 32; DB 15; Length 54;  
Best Local Similarity 100.0%; Pred. No. 3.3e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGPCV 6  
DB 38 GGPCV 42

RESULT 38  
US-10-424-599-152102  
Sequence 152102, Application US/10424599  
Publication No. US20040031072A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa Thomas J  
APPLICANT: Kovalic David K  
APPLICANT: Zhou Yihua  
APPLICANT: Cao Yongwei  
TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21(53223)B  
CURRENT APPLICATION NUMBER: US/10/424,599  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 285684  
SEQ ID NO 152102  
LENGTH: 57  
TYPE: PRT  
ORGANISM: Glycine max  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (1) (57)  
OTHER INFORMATION: unsure at all Xaa locations  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT3847\_10837C.1.pep  
US-10-424-599-152102

Query Match 88.9%; Score 32; DB 15; Length 57;  
Best Local Similarity 100.0%; Pred. No. 3.4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGPCV 6  
DB 20 GGPCV 24

RESULT 39  
US-10-425-115-336645  
Sequence 336645, Application US/10425115  
Publication No. US20040214272A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa Thomas J  
APPLICANT: Kovalic David K  
APPLICANT: Zhou Yihua  
APPLICANT: Cao Yongwei  
TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
TITLE OF INVENTION: Plants  
FILE REFERENCE: 38-21(53222)B  
CURRENT APPLICATION NUMBER: US/10/425,115  
CURRENT FILING DATE: 2003-04-28  
NUMBER OF SEQ ID NOS: 369326  
SEQ ID NO 336645  
LENGTH: 57  
TYPE: PRT  
ORGANISM: Zea mays  
FEATURE:  
OTHER INFORMATION: Clone ID: MRT4577\_70149C.1.pep  
US-10-425-115-336645

Query Match 88.9%; Score 32; DB 17; Length 57;  
Best Local Similarity 100.0%; Pred. No. 3.4e+02;  
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AGPCV 5

Db 48 AGPC 52

RESULT 40

US-10-425-115-259867  
 ; Sequence 259867, Application US/10425115  
 ; Publication No. US20040214272A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: La Rosa, Thomas J.  
 ; APPLICANT: Kovalic, David K.  
 ; APPLICANT: Zhou, Yihua  
 ; APPLICANT: Cao, Yongwei  
 ; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With  
 ; TITLE OF INVENTION: Plants  
 ; FILE REFERENCE: 38-21(53222)B  
 ; CURRENT APPLICATION NUMBER: US/10/425,115  
 ; CURRENT FILING DATE: 2003-04-28  
 ; NUMBER OF SEQ ID NOS: 369326  
 ; SEQ ID NO 259867  
 ; LENGTH: 62  
 ; TYPE: PRT  
 ; ORGANISM: Zea mays  
 ; FEATURE:  
 ; OTHER INFORMATION: Clone ID: MRT4577\_168617C.1.pep  
 US-10-425-115-259867

Query Match 88.9%; Score 32; DB 17; Length 62;  
 Best Local Similarity 100.0%; Pred. No. 3.7e+02;  
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 AGPC 5  
 Db 36 AGPC 40

Search completed: December 30, 2004, 15:31:16  
 Job time : 4.73678 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: December 30, 2004, 15:14:08 : Search time 215.263 Seconds

(without alignments)  
2929.449 Million cell updates/sec

Title: US-10-719-385-2

Perfect score: 9007  
Sequence: 1 MIRSKRITSVLSPCRSSREL.....PEGEPLILQVQAFVRHMR 1753Scoring table: BLOSUM62  
Gap: 10.0, Gapext 0.5

Searched: 1599051 seqs, 359727711 residues

Total number of hits satisfying chosen parameters: 1599051

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 150 summaries

## Database :

Published Applications AA:\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US09C\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US10\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/1/pubpaa/US10D\_PUBCOMB.pep.\*
- 17: /cgn2\_6/ptodata/1/pubpaa/US10D\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/1/pubpaa/US11\_NEW\_PUB.pep.\*
- 19: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*
- 20: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	9007	100.0	1753	US-10-719-385-2	Sequence 2, Appli
2	9005	100.0	1753	US-10-719-385-4	Sequence 4, Appli
3	9005	100.0	1753	US-10-719-385-8	Sequence 8, Appli
4	9004	100.0	1753	US-10-719-385-3	Sequence 3, Appli
5	9004	100.0	1753	US-10-719-385-11	Sequence 11, Appli
6	9004	100.0	1753	US-10-719-385-13	Sequence 13, Appli
7	9004	100.0	1753	US-10-719-385-14	Sequence 14, Appli
8	9003	100.0	1753	US-10-719-385-15	Sequence 15, Appli
9	9002	99.9	1753	US-10-719-385-6	Sequence 6, Appli
10	9002	99.9	1753	US-10-719-385-10	Sequence 10, Appli
11	9001	99.9	1753	US-10-719-385-16	Sequence 16, Appli
12	8999	99.9	1753	US-10-719-385-7	Sequence 7, Appli
13	8999	99.9	1753	US-10-719-385-12	Sequence 12, Appli

14	8998	99.9	1753	US-10-719-385-18	Sequence 18, Appli
15	8997	99.9	1753	US-10-719-385-9	Sequence 9, Appli
16	8993	99.8	1753	US-10-719-385-17	Sequence 17, Appli
17	8991	99.8	1753	US-10-719-385-19	Sequence 19, Appli
18	8985.5	99.8	1752	US-10-719-385-5	Sequence 5, Appli
19	8939	99.2	1745	US-10-719-385-21	Sequence 21, Appli
20	8939	99.2	1745	US-10-370-115B-544	Sequence 544, Appli
21	8927	99.1	1745	US-10-719-385-22	Sequence 22, Appli
22	5270	58.5	1111	US-10-719-385-23	Sequence 23, Appli
23	4093	45.4	853	US-10-719-385-24	Sequence 24, Appli
24	2420	26.9	525	US-10-719-385-25	Sequence 25, Appli
25	369.5	4.1	2026	US-10-437-963-135530	Sequence 135530, Appli
26	305	3.4	63	US-10-106-698-5917	Sequence 5917, Appli
27	245	2.7	1745	US-10-437-963-173637	Sequence 173637, Appli
28	176.5	2.0	2905	US-10-437-963-154118	Sequence 154118, Appli
29	175.5	1.9	1676	US-10-128-714-8246	Sequence 8246, Appli
30	169.5	1.9	2621	US-10-437-963-122168	Sequence 122168, Appli
31	166	1.8	2122	US-10-437-963-189782	Sequence 189782, Appli
32	165	1.8	1545	US-10-128-714-3246	Sequence 3246, Appli
33	160.5	1.8	2462	US-10-437-963-114113	Sequence 114113, Appli
34	157	1.7	3859	US-10-408-765A-354	Sequence 354, Appli
35	156	1.7	2627	US-10-424-599-256710	Sequence 256710, Appli
36	155	1.7	1557	US-10-369-493-2224	Sequence 2224, Appli
37	155	1.7	2834	US-10-424-599-256711	Sequence 256711, Appli
38	153	1.7	3830	US-10-723-860-2568	Sequence 2568, Appli
39	152.5	1.7	2159	US-10-437-963-108860	Sequence 108860, Appli
40	152.5	1.7	2412	US-10-408-765A-214	Sequence 214, Appli
41	151	1.7	3225	US-10-408-765A-254	Sequence 254, Appli
42	151	1.7	4464	US-10-369-493-5019	Sequence 5019, Appli
43	148.5	1.6	2811	US-10-810-352-39	Sequence 39, Appli
44	148	1.6	2593	US-10-437-963-114115	Sequence 114115, Appli
45	147.5	1.6	2209	US-09-902-941-1903	Sequence 1903, Appli
46	147.5	1.6	2209	US-09-849-626-1903	Sequence 1903, Appli
47	147.5	1.6	2209	US-10-017-754-1903	Sequence 1903, Appli
48	147.5	1.6	2209	US-10-113-872-1903	Sequence 1903, Appli
49	147.5	1.6	2209	US-10-283-017-1903	Sequence 1903, Appli
50	147.5	1.6	2209	US-10-408-765A-863	Sequence 863, Appli
51	147.5	1.6	2209	US-10-723-860-1675	Sequence 1675, Appli
52	147.5	1.6	2859	US-10-087-192-249	Sequence 249, Appli
53	147	1.6	3907	US-10-171-111-2	Sequence 2, Appli
54	147	1.6	3925	US-10-171-111-6	Sequence 6, Appli
55	145.5	1.6	2848	US-10-369-493-6048	Sequence 6048, Appli
56	145	1.6	2311	US-10-739-930-10532	Sequence 10532, Appli
57	144.5	1.6	2549	US-09-950-634-3	Sequence 3, Appli
58	144.5	1.6	2549	US-10-688-016-3	Sequence 3, Appli
59	143.5	1.6	2405	US-10-437-963-106615	Sequence 106615, Appli
60	143.5	1.6	876	US-09-738-626-4789	Sequence 4789, Appli
61	143	1.6	1783	US-10-188-832-86	Sequence 86, Appli
62	143	1.6	2549	US-10-701-490-2	Sequence 2, Appli
63	143	1.6	3899	US-10-171-111-4	Sequence 4, Appli
64	143	1.6	3917	US-10-171-111-8	Sequence 8, Appli
65	142.5	1.6	973	US-10-437-963-178286	Sequence 178286, Appli
66	142.5	1.6	2545	US-10-092-900A-76	Sequence 76, Appli
67	142	1.6	2646	US-10-437-963-189923	Sequence 189923, Appli
68	140.5	1.6	1821	US-10-437-963-203003	Sequence 203003, Appli
69	140.5	1.6	2852	US-10-481-471-10	Sequence 10, Appli
70	140	1.6	1867	US-10-437-963-137806	Sequence 137806, Appli
71	140	1.6	3878	US-10-080-608A-11	Sequence 11, Appli
72	140	1.6	3911	US-10-370-685-100	Sequence 100, Appli
73	140	1.6	3911	US-10-408-765A-1839	Sequence 1839, Appli
74	139.5	1.5	4096	US-10-473-127-571	Sequence 571, Appli
75	139.5	1.5	4097	US-10-363-616-415	Sequence 415, Appli
76	139.5	1.5	4097	US-10-408-765A-1558	Sequence 1558, Appli
77	139.5	1.5	4097	US-10-473-127-569	Sequence 569, Appli
78	139.5	1.5	4128	US-10-363-616-416	Sequence 416, Appli
79	139.5	1.5	4128	US-10-473-127-574	Sequence 574, Appli
80	138.5	1.5	1651	US-10-128-558-28	Sequence 28, Appli
81	138.5	1.5	2132	US-10-424-599-161362	Sequence 161362, Appli
82	138.5	1.5	2835	US-09-885-535-4	Sequence 4, Appli
83	138.5	1.5	4095	US-10-473-127-572	Sequence 572, Appli
84	138	1.5	1479	US-10-437-963-106122	Sequence 106122, Appli
85	136.5	1.5	4128	US-10-473-127-568	Sequence 568, Appli
86	136.5	1.5	4128	US-10-473-127-573	Sequence 573, Appli

87	135.5	1.5	2335	14	US-10-369-493-2561	Sequence 2561, Ap
88	135	1.5	1230	14	US-10-205-219-44	Sequence 44, Appl
89	135	1.5	1230	14	US-10-205-219-97	Sequence 97, Appl
90	134.5	1.5	1417	16	US-10-408-765A-671	Sequence 671, App
91	134.5	1.5	2222	14	US-10-369-493-3923	Sequence 3923, Ap
92	134	1.5	1905	15	US-10-259-194A-86	Sequence 86, Appl
93	134	1.5	2470	14	US-10-369-493-22191	Sequence 22191, A
94	134	1.5	5183	14	US-10-107-521-1	Sequence 1, Appl1
95	133.5	1.5	789	16	US-10-437-963-119795	Sequence 1, Appl1
96	133.5	1.5	1088	13	US-10-001-867-127	Sequence 127, App
97	133.5	1.5	1088	14	US-10-074-475-255	Sequence 255, App
98	133	1.5	1494	14	US-10-195-144-17	Sequence 17, Appl
99	133	1.5	1494	14	US-10-345-072-17	Sequence 17, Appl
100	132.5	1.5	1837	14	US-10-369-493-22734	Sequence 22734, A
101	131.5	1.5	2055	13	US-10-017-216-4	Sequence 4, Appl1
102	131.5	1.5	2151	15	US-10-424-599-161366	Sequence 161366,
103	131	1.5	2221	15	US-10-424-599-161365	Sequence 161365,
104	131	1.5	2677	16	US-10-367-094-124	Sequence 124, App
105	131	1.5	4128	14	US-10-205-194-1	Sequence 1, Appl1
106	130	1.4	1819	16	US-10-437-963-111842	Sequence 111842,
107	130	1.4	2111	16	US-10-437-963-125597	Sequence 125597,
108	129.5	1.4	1818	14	US-10-032-585-7616	Sequence 7616, Ap
109	129.5	1.4	1826	15	US-10-289-762-113	Sequence 113, App
110	129	1.4	1966	14	US-10-369-493-6049	Sequence 6049, Ap
111	128.5	1.4	1826	15	US-10-282-122A-54869	Sequence 54869, A
112	128.5	1.4	3092	9	US-09-801-369-172	Sequence 172, App
113	128.5	1.4	3092	14	US-10-369-493-1470	Sequence 1470, Ap
114	128	1.4	1166	17	US-10-370-715B-562	Sequence 562, App
115	128	1.4	2458	14	US-10-231-172-257	Sequence 257, App
116	128	1.4	2458	15	US-10-221-278-257	Sequence 257, App
117	128	1.4	3006	15	US-10-210-172-36	Sequence 36, Appl
118	127.5	1.4	1960	15	US-10-313-203-27	Sequence 27, Appl
119	127.5	1.4	2517	15	US-10-313-203-2	Sequence 2, Appl1
120	127.5	1.4	3433	16	US-10-408-765A-731	Sequence 731, App
121	127.5	1.4	3433	16	US-10-408-765A-732	Sequence 732, App
122	127.5	1.4	3829	16	US-10-408-765A-849	Sequence 849, App
123	127.5	1.4	3829	17	US-10-723-860-2137	Sequence 2137, App
124	127.5	1.4	3863	15	US-10-197-824-7	Sequence 7, Appl1
125	127.5	1.4	5171	16	US-10-408-765A-2687	Sequence 2687, Ap
126	126.5	1.4	1542	16	US-10-408-765A-269	Sequence 269, App
127	126.5	1.4	2671	16	US-10-408-765A-442	Sequence 442, App
128	126	1.4	808	14	US-10-104-047-2324	Sequence 2324, Ap
129	126	1.4	925	14	US-10-369-493-2281	Sequence 2281, Ap
130	126	1.4	4025	16	US-10-437-963-193926	Sequence 193926,
131	126	1.4	5373	16	US-10-408-765A-741	Sequence 741, App
132	125.5	1.4	920	16	US-10-437-963-136710	Sequence 136710,
133	125.5	1.4	1377	14	US-10-205-342-25	Sequence 25, Appl
134	125.5	1.4	1464	17	US-10-331-053-83	Sequence 83, Appl
135	125.5	1.4	2541	14	US-10-177-293-470	Sequence 470, App
136	125	1.4	1933	17	US-10-723-860-1971	Sequence 1971, App
137	124.5	1.4	916	16	US-10-473-574-22	Sequence 22, Appl
138	124.5	1.4	1475	14	US-10-007-047-2	Sequence 2, Appl1
139	124	1.4	821	14	US-10-369-493-2791	Sequence 2791, Ap
140	124	1.4	1209	15	US-10-282-122A-67611	Sequence 67611, A
141	124	1.4	2457	16	US-10-437-963-115622	Sequence 115622,
142	124	1.4	4313	15	US-10-287-226-312	Sequence 312, App
143	123.5	1.4	1293	16	US-10-437-963-185505	Sequence 185505,
144	123.5	1.4	1293	16	US-10-007-047-4	Sequence 4, Appl1
145	123.5	1.4	1451	14	US-09-998-027-1	Sequence 1, Appl1
146	123.5	1.4	1451	14	US-10-165-099-1	Sequence 1, Appl1
147	123.5	1.4	1471	10	US-09-998-027-4	Sequence 4, Appl1
148	123.5	1.4	1471	11	US-10-165-099-4	Sequence 4, Appl1
149	123.5	1.4	2053	14	US-09-964-956-11	Sequence 11, Appl
150	123.5	1.4	2053	15	US-10-262-511-2	Sequence 2, Appl1

## ALIGNMENTS

RESULT 1  
US-10-719-385-2  
; Sequence 2, Application US/10719385  
; Publication No. US20040209284A1

GENERAL INFORMATION:									
; APPLICANT: O'Toole et al.									
; TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis									
; FILE REFERENCE: 22058-582									
; CURRENT APPLICATION NUMBER: US/10/719,385									
; CURRENT FILING DATE: 2003-11-21									
; PRIOR APPLICATION NUMBER: PCT/US03/37339									
; PRIOR FILING DATE: 2003-11-21									
; PRIOR APPLICATION NUMBER: 60/428,094									
; PRIOR FILING DATE: 2002-11-21									
; NUMBER OF SEQ ID NOS: 26									
; SOFTWARE: PatentIn Ver. 2.1									
; SEQ ID NO 2									
; LENGTH: 1753									
; TYPE: PRT									
; ORGANISM: Homo sapiens									
US-10-719-385-2									
Query Match 100.0%; Score 9007; DB 17; Length 1753;									
Best Local Similarity 100.0%; Pred. No. 0;									
Matches 1753; Conservative 0; Mismatches 0; Indels 0; Gaps 0;									
Qy	1	MIRSKITSVLSFCRSSRELMTLLGRSALRELQTEAEINLKMRRLLLEGLSYKPPSPS	60						
Db	1	MIRSKITSVLSFCRSSRELMTLLGRSALRELQTEAEINLKMRRLLLEGLSYKPPSPS	60						
Qy	61	SAEVKANKVAPLKLGLRISKFLGLDEQSVOLLOCTLOEDYRGTROSVKTVLQDER	120						
Db	61	SAEVKANKVAPLKLGLRISKFLGLDEQSVOLLOCTLOEDYRGTROSVKTVLQDER	120						
Qy	121	QSQALIKLADYYEERTCLRCVHLITLTFQDERHPHYREYVADCVKLEKELVSKYRQ	180						
Db	121	QSQALIKLADYYEERTCLRCVHLITLTFQDERHPHYREYVADCVKLEKELVSKYRQ	180						
Qy	181	PEELVYKTEAPTWETHGNLMTEROVSRWFVQCLREQSMLEIIFLYAYAFEMASDILVLT	240						
Db	181	PEELVYKTEAPTWETHGNLMTEROVSRWFVQCLREQSMLEIIFLYAYAFEMASDILVLT	240						
Qy	241	KMFEEQFSGRQNRHLVDETMDPFVDRIGYFSLILVEQMDIESLHKALDREHOF	300						
Db	241	KMFEEQFSGRQNRHLVDETMDPFVDRIGYFSLILVEQMDIESLHKALDREHOF	300						
Qy	301	AQGLICQDMDCLMTFGDIPHNAVYLALMRLHTLNPEETSVAVKIGTALQNLVFO	360						
Db	301	AQGLICQDMDCLMTFGDIPHNAVYLALMRLHTLNPEETSVAVKIGTALQNLVFO	360						
Qy	361	YLRRLQSLASGNDCTTACMCVYGLSFVYTSLEHLTNGQODIIDTACEVLADPSL	420						
Db	361	YLRRLQSLASGNDCTTACMCVYGLSFVYTSLEHLTNGQODIIDTACEVLADPSL	420						
Qy	421	PELFWGTEPISGILIDSVCQMFPHLSPLQLLALVSGKSTAKKVSFLLKMSFVNE	480						
Db	421	PELFWGTEPISGILIDSVCQMFPHLSPLQLLALVSGKSTAKKVSFLLKMSFVNE	480						
Qy	481	LYHKHPDIVSHDGTLMRRQTPKLLYPLGGQTNLRIPOGTQVQVMDLDRALVWREYSY	540						
Db	481	LYHKHPDIVSHDGTLMRRQTPKLLYPLGGQTNLRIPOGTQVQVMDLDRALVWREYSY	540						
Qy	541	SSMTLFCLEMLLHVSTADVQHCORVPIIDLVHXYSTDLSTADCLPTTSRYML	600						
Db	541	SSMTLFCLEMLLHVSTADVQHCORVPIIDLVHXYSTDLSTADCLPTTSRYML	600						
Qy	601	LQRLTVISPPVIVASCVNCLFVLAAAPPAKWTDLRHGFLPFAHPVSSLSQMTSAE	660						
Db	601	LQRLTVISPPVIVASCVNCLFVLAAAPPAKWTDLRHGFLPFAHPVSSLSQMTSAE	660						
Qy	661	GNAGGYGNLNNSEPOGEYGVITAFRLITTLVKGQSGTOSQGLVPCVWFVKEMLP	720						
Db	661	GNAGGYGNLNNSEPOGEYGVITAFRLITTLVKGQSGTOSQGLVPCVWFVKEMLP	720						
Qy	721	SYHKRWYNSHGVEQIGCLILELHAILNLCHETDLHSSHTPSLQFLCICSLAYTEAGOT	780						
Db	721	SYHKRWYNSHGVEQIGCLILELHAILNLCHETDLHSSHTPSLQFLCICSLAYTEAGOT	780						



781 VININGICVDITDMMAOPRSDGAGGQGLIKITVLAFTSYNNVIRLKPENNVSP 840  
781 VININGICVDITDMMAOPRSDGAGGQGLIKITVLAFTSYNNVIRLKPENNVSP 840  
841 LEQALSGHAGNNLIALVAKTIYKHPALPRLAIOQLKRLATVAPMSVYACLGNDAA 900  
841 LEQALSGHAGNNLIALVAKTIYKHPALPRLAIOQLKRLATVAPMSVYACLGNDAA 900  
901 IRDAFLTLQSKIEDMRKIMWILEFLVAVETOPGLIELFLNLEVKDSDGSKERSLGMW 960  
901 IRDAFLTLQSKIEDMRKIMWILEFLVAVETOPGLIELFLNLEVKDSDGSKERSLGMW 960  
961 SCILAVLILISQODRWCPPLHRAIAFLHLMORRDSAMVLTATKPEMNLSP 1020  
961 SCILAVLILISQODRWCPPLHRAIAFLHLMORRDSAMVLTATKPEMNLSP 1020  
1021 LFGTLPSESESEPIETCALIMKICLEIYVVKSGIDSLKDTLKKFSIEKFPAYMS 1080  
1021 LFGTLPSESESEPIETCALIMKICLEIYVVKSGIDSLKDTLKKFSIEKFPAYMS 1080  
1081 GYVKSIAVHVAETEGSSCTSLLEYQMLVSAMRMLIATTHADIMHLDTSVVRQLFLDV 1140  
1081 GYVKSIAVHVAETEGSSCTSLLEYQMLVSAMRMLIATTHADIMHLDTSVVRQLFLDV 1140  
1141 LDGTALLVPAASVNCIRLGSWKTLLILILRQMRKELGSVDEILGPILEIEGLQADQ 1200  
1141 LDGTALLVPAASVNCIRLGSWKTLLILILRQMRKELGSVDEILGPILEIEGLQADQ 1200  
1201 QLMETKAKVSAFTVLOMKEMKYSIDIPQYSQVLVANCETLOEVIALLPOTRSLALG 1260  
1201 QLMETKAKVSAFTVLOMKEMKYSIDIPQYSQVLVANCETLOEVIALLPOTRSLALG 1260  
1261 SATEDKDSMETDDCSRSRHRDQDVCVGLHLAVELCEVEDGDSWLOVTRRLPILPTL 1320  
1261 SATEDKDSMETDDCSRSRHRDQDVCVGLHLAVELCEVEDGDSWLOVTRRLPILPTL 1320  
1321 LTTLEVSIRMKONHAFTEATLHLLTLARTOOGATVAVAGITOSTICPLSVYOLSTNG 1380  
1321 LTTLEVSIRMKONHAFTEATLHLLTLARTOOGATVAVAGITOSTICPLSVYOLSTNG 1380  
1381 TAQTSASRKSIDASMPGCVYRLSMSLEQLKTLRYNPLPALDPVGHOBERTLOCLNA 1440  
1381 TAQTSASRKSIDASMPGCVYRLSMSLEQLKTLRYNPLPALDPVGHOBERTLOCLNA 1440  
1441 VRTVQSLACLEBADHTVGFIIQLSNFMKEWHFLPQLMRDIQVNLGYLCOACTSLHSRK 1500  
1441 VRTVQSLACLEBADHTVGFIIQLSNFMKEWHFLPQLMRDIQVNLGYLCOACTSLHSRK 1500  
1501 MLOHVLQKNGDGLSAAVAVQVQRPSPASAPSSSKOPADTEASEOQALHTVOYGLK 1560  
1501 MLOHVLQKNGDGLSAAVAVQVQRPSPASAPSSSKOPADTEASEOQALHTVOYGLK 1560  
1561 ILSTKLALRHFTPVCOILIDQSLDLAEYNFLPALSTFTPTPDESEVAPSGFTLLATVNV 1620  
1561 ILSTKLALRHFTPVCOILIDQSLDLAEYNFLPALSTFTPTPDESEVAPSGFTLLATVNV 1620  
1621 ALNMLGELDKKKEPLTOAVGLSTOAGRTILKSLMFTMENCFYLLISQAMRYLDPVAVH 1680  
1621 ALNMLGELDKKKEPLTOAVGLSTOAGRTILKSLMFTMENCFYLLISQAMRYLDPVAVH 1680  
1681 PRDKRMQOELSESLSTLSSLSRYFRGAPSSPATGVLPSPQGSTLSISKASPPSOEPL 1740  
1681 PRDKRMQOELSESLSTLSSLSRYFRGAPSSPATGVLPSPQGSTLSISKASPPSOEPL 1740  
1741 IQLVQAFVRHMQR 1753  
1741 IQLVQAFVRHMQR 1753

RESULT 2  
US-10-719-385-4  
; Sequence 4, Application US/10719385

Publication No. US20040209284A1  
GENERAL INFORMATION:  
APPLICANT: O'Toole et al.  
TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis  
FILE REFERENCE: 22058-582  
CURRENT APPLICATION NUMBER: US/10/719,385  
CURRENT FILING DATE: 2003-11-21  
PRIOR APPLICATION NUMBER: PCT/US03/37339  
PRIOR FILING DATE: 2003-11-21  
PRIOR APPLICATION NUMBER: 60/428,094  
PRIOR FILING DATE: 2002-11-21  
NUMBER OF SEQ ID NOS: 26  
SOFTWARE: Patent In Ver. 2.1  
SEQ ID NO 4  
LENGTH: 1753  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-719-385-4  
Query Match 100.0%; Score 9005; DB 17; Length 1753;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1752; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
1 MIRSKITSVLSFCSSRHEMTIILGRSALRELQIEALNKHWRRLLEGSIYKPPSPS 60  
1 MIRSKITSVLSFCSSRHEMTIILGRSALRELQIEALNKHWRRLLEGSIYKPPSPS 60  
61 SAEKVKANKDVASPLKEIGLRISKFLGLDEBQSVOLLOCCYIOEDYGRTRDSYKTVLODER 120  
61 SAEKVKANKDVASPLKEIGLRISKFLGLDEBQSVOLLOCCYIOEDYGRTRDSYKTVLODER 120  
121 OSQALILKADYVVEERFCILRCVHLITVRODEHPRVAVADCVDLREKIVSKYQO 180  
121 OSQALILKADYVVEERFCILRCVHLITVRODEHPRVAVADCVDLREKIVSKYQO 180  
181 FEELYKTEAPWETHGNLMTERQVSRWFVQCLREOSMLLEIIFLYAYFEMAPSDLVLT 240  
181 FEELYKTEAPWETHGNLMTERQVSRWFVQCLREOSMLLEIIFLYAYFEMAPSDLVLT 240  
241 KMFKEQGSQOTNRHLVDETMDFVDRIGYFSALILVEGMDIESLHCALDDRRREHOF 300  
241 KMFKEQGSQOTNRHLVDETMDFVDRIGYFSALILVEGMDIESLHCALDDRRREHOF 300  
301 AODGLICQDMCMLTFGDIPIHNAVLLAMLLRRTLNPEETSSVVRKIGTAIOLNVQ 360  
301 AODGLICQDMCMLTFGDIPIHNAVLLAMLLRRTLNPEETSSVVRKIGTAIOLNVQ 360  
361 YLTRLQSLAGSNDCTSTACMCYGLSFVLSLELHTLGNQODIIDTACEVLADPSL 420  
361 YLTRLQSLAGSNDCTSTACMCYGLSFVLSLELHTLGNQODIIDTACEVLADPSL 420  
421 PELFWGTEPTSGLIILDSVCGMPHLLSPILQILRALVSGSTAKKYSPFLDKMSFYNE 480  
421 PELFWGTEPTSGLIILDSVCGMPHLLSPILQILRALVSGSTAKKYSPFLDKMSFYNE 480  
481 LYKXKPHVISHEDGTLMRQTPKLLYPLGGQTMRIQGVVGMULDDRAVLVWEXSY 540  
481 LYKXKPHVISHEDGTLMRQTPKLLYPLGGQTMRIQGVVGMULDDRAVLVWEXSY 540  
541 SSWTLFCEIEMLLHVTADVIOHCORVKPILIDLVHVIETDLSIADCLPITRIYML 600  
541 SSWTLFCEIEMLLHVTADVIOHCORVKPILIDLVHVIETDLSIADCLPITRIYML 600  
601 LORLTTVISPVDVIAVCVNCITVLAARNPAKWMTDLRHTGFLPFAHVPSSLSQMSIAE 660  
601 LORLTTVISPVDVIAVCVNCITVLAARNPAKWMTDLRHTGFLPFAHVPSSLSQMSIAE 660  
661 GNNAGGYNLLMNSPOGEGYGVTAFLRLITTLVKGQSGTOSQGLVPCVMFVLKMLP 720  
661 GNNAGGYNLLMNSPOGEGYGVTAFLRLITTLVKGQSGTOSQGLVPCVMFVLKMLP 720  
721 SYHKRVYNSHGREOIGCLIELLHAILNLGHEITLHSSHTPSLOFLCISLAVTEAGOT 780  
721 SYHKRVYNSHGREOIGCLIELLHAILNLGHEITLHSSHTPSLOFLCISLAVTEAGOT 780

721 SYHKRNSHGVREIQICLILIELIAIINLCHETDLSHSHHTPSLOFLCISLATEAGOT 780  
781 VININGIGVDTIDVMAAOPRSDGAEQGGQGLIKTYKAFSVNNNTIRLKPENVSP 840  
781 VININGIGVDTIDVMAAOPRSDGAEQGGQGLIKTYKAFSVNNNTIRLKPENVSP 840  
841 LEQALSOGAHGNLJAVLAKYIYKHDPALPRALIQLLKRLATVAPSVYACLGNDAAA 900  
841 LEQALSOGAHGNLJAVLAKYIYKHDPALPRALIQLLKRLATVAPSVYACLGNDAAA 900  
901 IRDAFLTLQSKIEMRIKMLLEFLTVAVETQPGILFNLKXKSDSDSKESLGMW 960  
901 IRDAFLTLQSKIEMRIKMLLEFLTVAVETQPGILFNLKXKSDSDSKESLGMW 960  
961 SCIAHVLILDSQODRWCPPLHRAAIAFLHAIAMODRBSAMVLRKPKFMENTLSP 1020  
961 SCIAHVLILDSQODRWCPPLHRAAIAFLHAIAMODRBSAMVLRKPKFMENTLSP 1020  
1021 LFGTSPSESTSEPSILETCALIMKICLIEIYVVGSLDOSLKDOLKKSIEKRFAYMS 1080  
1021 LFGTSPSESTSEPSILETCALIMKICLIEIYVVGSLDOSLKDOLKKSIEKRFAYMS 1080  
1081 GYVSLAHVAETESSCTSLLEYQMLVSAWMLIITTHADIMHLDTSVVRQLFLDV 1140  
1081 GYVSLAHVAETESSCTSLLEYQMLVSAWMLIITTHADIMHLDTSVVRQLFLDV 1140  
1141 LDGTALLVPAVAVNCLRLGSMKCTLLILRLQWRELGSDVEILGPTTEILEGLODQ 1200  
1141 LDGTALLVPAVAVNCLRLGSMKCTLLILRLQWRELGSDVEILGPTTEILEGLODQ 1200  
1201 QLMETKAKVPSAFTVLQMKEMKVSIPQYSQVLANVCELTQOEVIALFDQTHSIALG 1260  
1201 QLMETKAKVPSAFTVLQMKEMKVSIPQYSQVLANVCELTQOEVIALFDQTHSIALG 1260  
1261 SATEDKSMETDSCRSRHRDQDGVCLGHLAKELCEVDEDEGDSWQVTRRLPIPLET 1320  
1261 SATEDKSMETDSCRSRHRDQDGVCLGHLAKELCEVDEDEGDSWQVTRRLPIPLET 1320  
1321 LTTLEVSILRMKONLHTEATLHLLTLTARQOGATAVAGAGITQSIQPLSVYQLSITNG 1380  
1321 LTTLEVSILRMKONLHTEATLHLLTLTARQOGATAVAGAGITQSIQPLSVYQLSITNG 1380  
1381 TAQTPSARSKSLDAPSWGCVYRLSMLBOLKTRNPLPEALDFGCVHQBERTLOCLNA 1440  
1381 TAQTPSARSKSLDAPSWGCVYRLSMLBOLKTRNPLPEALDFGCVHQBERTLOCLNA 1440  
1441 VRTVQSLACLBEADHTVGFILQLSNFMKEMHFLPOLMRDIOVNLGYLCOACTSLHSRK 1500  
1441 VRTVQSLACLBEADHTVGFILQLSNFMKEMHFLPOLMRDIOVNLGYLCOACTSLHSRK 1500  
1501 MLQHYLQNKNGDGLPSAVAQEVORPPSASASAPSSKQPAADTEASEGOALHTVQYGLK 1560  
1501 MLQHYLQNKNGDGLPSAVAQEVORPPSASASAPSSKQPAADTEASEGOALHTVQYGLK 1560  
1561 ILSTLALHRTFTDVCQILLDOSIDLAEYNFLPALSTFTPTPSEVAPSGITLAVNV 1620  
1561 ILSTLALHRTFTDVCQILLDOSIDLAEYNFLPALSTFTPTPSEVAPSGITLAVNV 1620  
1621 ALNMLGELDKKKEPLTOAVGSLTOAEGRTLSLIMFTMENCFTLLISQAMRYLRDPAVH 1680  
1621 ALNMLGELDKKKEPLTOAVGSLTOAEGRTLSLIMFTMENCFTLLISQAMRYLRDPAVH 1680  
1681 PRDKQRMKQELSSSELSTLSSLSRYFRGABSSPATGVLPSPQKSTSLSKASPESEPL 1740  
1681 PRDKQRMKQELSSSELSTLSSLSRYFRGABSSPATGVLPSPQKSTSLSKASPESEPL 1740  
1741 IQLVQAFVRHMOR 1753  
1741 IQLVQAFVRHMOR 1753

Sequence 8, Application US/10719385  
Publication No. US20040209284A1  
GENERAL INFORMATION:  
APPLICANT: O'Toole et al.  
TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis  
FILE REFERENCE: 22058-582  
CURRENT APPLICATION NUMBER: US/10/719,385  
CURRENT FILING DATE: 2003-11-21  
PRIOR APPLICATION NUMBER: PCT/US03/37339  
PRIOR FILING DATE: 2003-11-21  
PRIOR APPLICATION NUMBER: 60/428,094  
PRIOR FILING DATE: 2002-11-21  
NUMBER OF SEQ ID NOS: 26  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO: 8  
LENGTH: 1753  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-10-719-385-8

Query Match 100.0%; Score 9005; DB 17; Length 1753;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1752; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

1 MIRSKITSVLSFCRSSREIMTLILGRSALRELSQIEAEINKMRRLLEGSLYKPPSPS 60  
1 MIRSKITSVLSFCRSSREIMTLILGRSALRELSQIEAEINKMRRLLEGSLYKPPSPS 60  
61 SAEYKANKVAPBKLGLRISKFLGLDEBSVOLIOCYLOEDYRGTDSVKTLODER 120  
61 SAEYKANKVAPBKLGLRISKFLGLDEBSVOLIOCYLOEDYRGTDSVKTLODER 120  
61 SAEYKANKVAPBKLGLRISKFLGLDEBSVOLIOCYLOEDYRGTDSVKTLODER 120  
121 QSQALIKIDYVYEEETCLRCVLHLLTFQOERHPYRYEYVDCVJLKEKLYSKYRQ 180  
121 QSQALIKIDYVYEEETCLRCVLHLLTFQOERHPYRYEYVDCVJLKEKLYSKYRQ 180  
121 QSQALIKIDYVYEEETCLRCVLHLLTFQOERHPYRYEYVDCVJLKEKLYSKYRQ 180  
181 FEEIYKTEADPTWETHGNLMTEROVSWMFVQCLREQSMLEIIFLYYAFEMAPSDLLVT 240  
181 FEEIYKTEADPTWETHGNLMTEROVSWMFVQCLREQSMLEIIFLYYAFEMAPSDLLVT 240  
181 FEEIYKTEADPTWETHGNLMTEROVSWMFVQCLREQSMLEIIFLYYAFEMAPSDLLVT 240  
241 KMFEEQFSGRQTRNRLVDETMDPVDRLGYFSALLIVEEMDIESLHKCALDRRELHQF 300  
241 KMFEEQFSGRQTRNRLVDETMDPVDRLGYFSALLIVEEMDIESLHKCALDRRELHQF 300  
301 AODGLICQDMDCLMTFEGDIPHHAPVLLAWALRHTLNPEETSVMRKIGTAIQLVNFQ 360  
301 AODGLICQDMDCLMTFEGDIPHHAPVLLAWALRHTLNPEETSVMRKIGTAIQLVNFQ 360  
361 YLTRLQSLASGNDCTTSTACMCVYGLSFVLSLELHTLGNQODIIDTACEVLADPSL 420  
361 YLTRLQSLASGNDCTTSTACMCVYGLSFVLSLELHTLGNQODIIDTACEVLADPSL 420  
421 PELFMGTEPISGIIILDSVCGMFPHLSPLOLALVSGSKTAKVVSFLDKMSFYNE 480  
421 PELFMGTEPISGIIILDSVCGMFPHLSPLOLALVSGSKTAKVVSFLDKMSFYNE 480  
481 LYHKKHDVISHEDGLTMRQTPKLLYPLGGQTNLRIPQSTVQWMLDPRALVREYSY 540  
481 LYHKKHDVISHEDGLTMRQTPKLLYPLGGQTNLRIPQSTVQWMLDPRALVREYSY 540  
541 SSMTLFTCEIEMILHVVSTADVIQHOCQVYPIIDLHXYISTDLSTADCLPITSRIYML 600  
541 SSMTLFTCEIEMILHVVSTADVIQHOCQVYPIIDLHXYISTDLSTADCLPITSRIYML 600  
601 LQRLTVISPPVIVIASCVNCLTVLAARPAKWMDLRHTGFLPFAHPVSSISQMSAE 660  
601 LQRLTVISPPVIVIASCVNCLTVLAARPAKWMDLRHTGFLPFAHPVSSISQMSAE 660  
661 GNNAGGYNLNMSQPOGEYGVITAFRLITTLVYGQISQSGQGLVPCVMEVLEMLP 720  
661 GNNAGGYNLNMSQPOGEYGVITAFRLITTLVYGQISQSGQGLVPCVMEVLEMLP 720  
721 SYHKRNSHGVREIQICLILIELIAIINLCHETDLSHSHHTPSLOFLCISLATEAGOT 780

Db 721 SYHKRYSHGREGIGCLILELHIAILNLCHETDLSHSHTPSLQFLICISLAATEAGCT 780  
Qy 781 VINIMGICVDITIDWMAAQPSRSDGEGGOGOLIKYTKALPSTYNNVIRLKPSPNVSP 840  
Db 781 VINIMGICVDITIDWMAAQPSRSDGEGGOGOLIKYTKALPSTYNNVIRLKPSPNVSP 840  
Qy 841 LEQALSOHGAGNNLIALAKYIYKHIOPALPRALITOLKRLATAPWVSACIGNDAAA 900  
Db 841 LEQALSOHGAGNNLIALAKYIYKHIOPALPRALITOLKRLATAPWVSACIGNDAAA 900  
Qy 901 IRDAFLTRLOSKIEMDMRIKMIWLEFLVAVETOPGLIELFNLLEVKGSDSKESLGMW 960  
Db 901 IRDAFLTRLOSKIEMDMRIKMIWLEFLVAVETOPGLIELFNLLEVKGSDSKESLGMW 960  
Qy 961 SCLHVAVELIISOQODRWCPPLHRAAIAFLHLMORRDSAMLVLTATKCPMENTLSP 1020  
Db 961 SCLHVAVELIISOQODRWCPPLHRAAIAFLHLMORRDSAMLVLTATKCPMENTLSP 1020  
Qy 1021 LFGTSPSESTSEPSILETCALIMKICLIEIYVVKGLSDSLKDTLKKFSIEKRFAYMS 1080  
Db 1021 LFGTSPSESTSEPSILETCALIMKICLIEIYVVKGLSDSLKDTLKKFSIEKRFAYMS 1080  
Qy 1081 GYVKSIAVHVAETEGSSCTSLLEYOMLVSAWRMLIATTHADIMHLDTSVVRQLFLDV 1140  
Db 1081 GYVKSIAVHVAETEGSSCTSLLEYOMLVSAWRMLIATTHADIMHLDTSVVRQLFLDV 1140  
Qy 1141 LDGKALLVPAVAVNCILGSMKCTLLILRLKMRKELSGVDEILGPIITELLEGVLQDQ 1200  
Db 1141 LDGKALLVPAVAVNCILGSMKCTLLILRLKMRKELSGVDEILGPIITELLEGVLQDQ 1200  
Qy 1201 QLMKTKYKVSFAFTVLQMKMKVSDIPOYSOLVANCETLQOEVIAPOTRHSALG 1260  
Db 1201 QLMKTKYKVSFAFTVLQMKMKVSDIPOYSOLVANCETLQOEVIAPOTRHSALG 1260  
Qy 1261 SATEDKDSMETDSCSRSHRDQDQCVLGLHLAKELCEVEDGDGSMLOVTRRLPIEPTL 1320  
Db 1261 SATEDKDSMETDSCSRSHRDQDQCVLGLHLAKELCEVEDGDGSMLOVTRRLPIEPTL 1320  
Qy 1321 LTTLEVSIRKMKONLHFTBATLHLITLARTQOGATVAGAGITQSICLPILSVYQSLTNG 1380  
Db 1321 LTTLEVSIRKMKONLHFTBATLHLITLARTQOGATVAGAGITQSICLPILSVYQSLTNG 1380  
Qy 1381 TAQPSASRSKSLDAPSWGVYRLSLSLEOLKTLRYNPLPEALDPFGVHOERTLQCLNA 1440  
Db 1381 TAQPSASRSKSLDAPSWGVYRLSLSLEOLKTLRYNPLPEALDPFGVHOERTLQCLNA 1440  
Qy 1441 VRTVQSLACLEADHTVGFIIQLSNFMKEMHFLPOLNRDIOVNLGYLCOACTSLHSRK 1500  
Db 1441 VRTVQSLACLEADHTVGFIIQLSNFMKEMHFLPOLNRDIOVNLGYLCOACTSLHSRK 1500  
Qy 1501 MLQHYLQNKNGDGLPSAVALQVORPPSAASAAPSSSKPPAADTEASEBOALHTVGYGLK 1560  
Db 1501 MLQHYLQNKNGDGLPSAVALQVORPPSAASAAPSSSKPPAADTEASEBOALHTVGYGLK 1560  
Qy 1561 ILSKTLALRHFTPTPVCOILLDOSIDLARINFLFALSTTTPTFDESEVAPSGTILATVY 1620  
Db 1561 ILSKTLALRHFTPTPVCOILLDOSIDLARINFLFALSTTTPTFDESEVAPSGTILATVY 1620  
Qy 1621 ALNMLGELDKKKEPLTQAVGLSTQAEGRITLSLMTMENC FYLLISQAMRYLRDPAVH 1680  
Db 1621 ALNMLGELDKKKEPLTQAVGLSTQAEGRITLSLMTMENC FYLLISQAMRYLRDPAVH 1680  
Qy 1681 PRDQKRMQOEISSEISTLSSLSRYFRGAPSSPATVLPSPQGSSTLSKASPSQSPPL 1740  
Db 1681 PRDQKRMQOEISSEISTLSSLSRYFRGAPSSPATVLPSPQGSSTLSKASPSQSPPL 1740  
Qy 1741 IQLVQAFVRHMOR 1753  
Db 1741 IQLVQAFVRHMOR 1753

RESULT 4

US-10-719-385-3  
; Sequence 3, Application US/10719385  
; Publication No. US20040209284A1  
; GENERAL INFORMATION:  
; APPLICANT: O'Toole et al.  
; TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis  
; FILE REFERENCE: 22058-582  
; CURRENT APPLICATION NUMBER: US/10/719,385  
; CURRENT FILING DATE: 2003-11-21  
; PRIOR APPLICATION NUMBER: PCT/US03/37339  
; PRIOR FILING DATE: 2003-11-21  
; PRIOR APPLICATION NUMBER: 60/428,094  
; PRIOR FILING DATE: 2002-11-21  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3  
; LENGTH: 1753  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-719-385-3

Query Match 100.0%; Score 9004; DB 17; Length 1753;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1752; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MIRSKITSVLSFCRSSRELWTILGRSALRELQIEAEIKMHRRLLEGSTYKPPSPS 60  
Db 1 MIRSKITSVLSFCRSSRELWTILGRSALRELQIEAEIKMHRRLLEGSTYKPPSPS 60  
Qy 61 SAEKVKANKDVAAPLKEIGLRISKFLGDEBQSVOLLQCYLOEDYRGTRDSVKTVLQDER 120  
Db 61 SAEKVKANKDVAAPLKEIGLRISKFLGDEBQSVOLLQCYLOEDYRGTRDSVKTVLQDER 120  
Qy 121 QSOALLIKIADYYEERTCIRCYLHLITFYQDERHRYRYEYACVDKLEKLSYKTRQO 180  
Db 121 QSOALLIKIADYYEERTCIRCYLHLITFYQDERHRYRYEYACVDKLEKLSYKTRQO 180  
Qy 181 FEEIYKTEAPTWETHNGNLMTEROYSRMFOCLREOSMLLEIIFLYYAYFEWAPSDLVLT 240  
Db 181 FEEIYKTEAPTWETHNGNLMTEROYSRMFOCLREOSMLLEIIFLYYAYFEWAPSDLVLT 240  
Qy 241 KMFEEQGSFQOTNRHLVDETMDFVDRIGYFSALILVEGNDIBSLHKCALDDBRRELHQF 300  
Db 241 KMFEEQGSFQOTNRHLVDETMDFVDRIGYFSALILVEGNDIBSLHKCALDDBRRELHQF 300  
Qy 301 AODGLICODMDCMLTREGDIPHHAPVLLAVALHHTLNPESTSSVVRKIGGTALQLVFQ 360  
Db 301 AODGLICODMDCMLTREGDIPHHAPVLLAVALHHTLNPESTSSVVRKIGGTALQLVFQ 360  
Qy 361 YLTELGLSISAGNDCTTSTACMCVYGLSFLVLSLELHTLGNQODIITRACEVLADPSL 420  
Db 361 YLTELGLSISAGNDCTTSTACMCVYGLSFLVLSLELHTLGNQODIITRACEVLADPSL 420  
Qy 421 PELFWGTEPTSGLGIILDSVCGMPPHLLSPILLQRLALVSGSKTAKKYYSPLDKMSFYNE 480  
Db 421 PELFWGTEPTSGLGIILDSVCGMPPHLLSPILLQRLALVSGSKTAKKYYSPLDKMSFYNE 480  
Qy 481 LYKXKPHDVISHBEGTLMRRQTPKLLPLPGQOTMLRIPQGVGQVMDLDRAYLVRWEYSY 540  
Db 481 LYKXKPHDVISHBEGTLMRRQTPKLLPLPGQOTMLRIPQGVGQVMDLDRAYLVRWEYSY 540  
Qy 541 SSMTLFCEIEMLHVSTADAVIOHCORVXPIIDLHVKVISTDISIADCLPITSRIYML 600  
Db 541 SSMTLFCEIEMLHVSTADAVIOHCORVXPIIDLHVKVISTDISIADCLPITSRIYML 600  
Qy 601 LQRLITTVISPPVDVIASCVNCLTVLAARNPAKWTDLRHTEFLPFVAHPVSSLQMSIAE 660  
Db 601 LQRLITTVISPPVDVIASCVNCLTVLAARNPAKWTDLRHTEFLPFVAHPVSSLQMSIAE 660  
Qy 661 GMAAGGYGNLIMNSQPOGEYGVITIAFLRLITTVKQGLSTQSGIIVPCVMFLKEMLP 720  
Db 661 GMAAGGYGNLIMNSQPOGEYGVITIAFLRLITTVKQGLSTQSGIIVPCVMFLKEMLP 720

QY 721 SYHKRYNSHGVREOIGCLIELIELIAINLCHETDLSHSHPSLOFLCISLAYTEAGOT 780  
 DB 721 SYHKRYNSHGVREOIGCLIELIELIAINLCHETDLSHSHPSLOFLCISLAYTEAGOT 780  
 QY 781 VININGIGVDTIDMMAAOPRSDGABGOGGOLIKTYKLAFSVTNNVIRLKPSPNVSP 840  
 DB 781 VININGIGVDTIDMMAAOPRSDGABGOGGOLIKTYKLAFSVTNNVIRLKPSPNVSP 840  
 QY 841 VININGIGVDTIDMMAAOPRSDGABGOGGOLIKTYKLAFSVTNNVIRLKPSPNVSP 840  
 DB 841 VININGIGVDTIDMMAAOPRSDGABGOGGOLIKTYKLAFSVTNNVIRLKPSPNVSP 840  
 QY 841 LEQALSOHGAHGNLJAVLAKYIYKHDPALPRLIOLKRLATAPAWSVACIGNDAA 900  
 DB 841 LEQALSOHGAHGNLJAVLAKYIYKHDPALPRLIOLKRLATAPAWSVACIGNDAA 900  
 QY 901 IRDAFLTRLOSKEIEMRIKVMILEFLVAVETOPGLIELFLNLEKSDSDSKESFLGMM 960  
 DB 901 IRDAFLTRLOSKEIEMRIKVMILEFLVAVETOPGLIELFLNLEKSDSDSKESFLGMM 960  
 QY 961 SCLHNVLELIDSOODRWCPPLHRAIAFLHAIAMORRBSANLVLTKEKMENTLSP 1020  
 DB 961 SCLHNVLELIDSOODRWCPPLHRAIAFLHAIAMORRBSANLVLTKEKMENTLSP 1020  
 QY 1021 LFGTLPSPSETSEPSILETCALIMKICLIEIYVYKSLDOSLKDOLTKKFSIEKRFAYWS 1080  
 DB 1021 LFGTLPSPSETSEPSILETCALIMKICLIEIYVYKSLDOSLKDOLTKKFSIEKRFAYWS 1080  
 QY 1081 GYVKSIAVHVAETEGSSCTSLLEYQMLVSAWRMLIATTHADIMHLDTSVVRQOLFV 1140  
 DB 1081 GYVKSIAVHVAETEGSSCTSLLEYQMLVSAWRMLIATTHADIMHLDTSVVRQOLFV 1140  
 QY 1141 LDGTALLVPAVNCLELGSMTCTLLILRLKQRELSGDEILGPTTELLEGVLQADQ 1200  
 DB 1141 LDGTALLVPAVNCLELGSMTCTLLILRLKQRELSGDEILGPTTELLEGVLQADQ 1200  
 QY 1201 QLMETKAKVSAFITTVLQMKEMKVDIPQYSQVLAVNCETLOEBVALFPQTHSLATG 1260  
 DB 1201 QLMETKAKVSAFITTVLQMKEMKVDIPQYSQVLAVNCETLOEBVALFPQTHSLATG 1260  
 QY 1261 SATEDKSMETDDCSRRHRDQDVCYLGLHAKELCEVDEBDGSDQVTRRLPIPTL 1320  
 DB 1261 SATEDKSMETDDCSRRHRDQDVCYLGLHAKELCEVDEBDGSDQVTRRLPIPTL 1320  
 QY 1321 LTTLEVSIRMKONHFTBATLHLTLTLARQOGATVAVAGITOSICLPLISVYQLSUNG 1380  
 DB 1321 LTTLEVSIRMKONHFTBATLHLTLTLARQOGATVAVAGITOSICLPLISVYQLSUNG 1380  
 QY 1381 TAQTPSASRKSJLDAPSWPGVYRLSMSLMEQLKTLRYNFLPEALDPVGHQERTLQCLNA 1440  
 DB 1381 TAQTPSASRKSJLDAPSWPGVYRLSMSLMEQLKTLRYNFLPEALDPVGHQERTLQCLNA 1440  
 QY 1441 VRTVQSLACLBEADHTVGFIIQLSNFMKEWHFHLPOLMRDIQVNLGYLCOACTSILHSRK 1500  
 DB 1441 VRTVQSLACLBEADHTVGFIIQLSNFMKEWHFHLPOLMRDIQVNLGYLCOACTSILHSRK 1500  
 QY 1501 MLQHYLQWKNKGDLPSAFAORVORPPSAASAAPSSKOPADTEASEQALHTVQYGLK 1560  
 DB 1501 MLQHYLQWKNKGDLPSAFAORVORPPSAASAAPSSKOPADTEASEQALHTVQYGLK 1560  
 QY 1561 ILSTKTLAALRHFTPDVQIILLDOSJDLAEYNFLALSTPTTPTPSEVAAPSEFTLLATVNY 1620  
 DB 1561 ILSTKTLAALRHFTPDVQIILLDOSJDLAEYNFLALSTPTTPTPSEVAAPSEFTLLATVNY 1620  
 QY 1621 ALNMLGELDKKKEPLTQAVGLSTQAEGRITKSLIMTMENCFLLLISQAMRYLADPAVH 1680  
 DB 1621 ALNMLGELDKKKEPLTQAVGLSTQAEGRITKSLIMTMENCFLLLISQAMRYLADPAVH 1680  
 QY 1681 PRDQRMKQELSELSTLSSLSRYFRRGAPSPATGVLPSPOKSTSLSKASPESEQPL 1740  
 DB 1681 PRDQRMKQELSELSTLSSLSRYFRRGAPSPATGVLPSPOKSTSLSKASPESEQPL 1740  
 QY 1741 IQLVQAFVRHQR 1753  
 DB 1741 IQLVQAFVRHQR 1753

RESULT 5  
 US-10-719-385-11  
 ; Sequence 11, Application US/10719385  
 ; Publication No. US20040209284A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: O'Toole et al.  
 ; TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis  
 ; FILE REFERENCE: 22058-582  
 ; CURRENT APPLICATION NUMBER: US/10/719,385  
 ; PRIOR APPLICATION NUMBER: PCT/US03/37339  
 ; PRIOR FILING DATE: 2003-11-21  
 ; PRIOR APPLICATION NUMBER: 60/428,094  
 ; NUMBER OF SEQ ID NOS: 26  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 11  
 ; LENGTH: 1753  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-10-719-385-11

Query Match 100.0%; Score 9004; DB 17; Length 1753;  
 Best Local Similarity 99.9%; Pred. No. 0;  
 Matches 1752; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MIRKSTSVLSFRCSSREIMTLLGRSALREISOIEAEINKHMRRLBGLSYKPPSPS 60  
 DB 1 MIRKSTSVLSFRCSSREIMTLLGRSALREISOIEAEINKHMRRLBGLSYKPPSPS 60  
 QY 61 SAEKRVANKOVAPLKEGLRISKFLGLDEQSOVLQCYLQBDYGRTRDSVTVLQDER 120  
 DB 61 SAEKRVANKOVAPLKEGLRISKFLGLDEQSOVLQCYLQBDYGRTRDSVTVLQDER 120  
 QY 121 QSOALLIKINDYYEERTCICRCVHLILTYFQDERHPYREYADCVDKLKEKYSKRRQ 180  
 DB 121 QSOALLIKINDYYEERTCICRCVHLILTYFQDERHPYREYADCVDKLKEKYSKRRQ 180  
 QY 181 FEELYTEAPWTETHGNLMTEROVSRMFVOCLEQSMLEIIFLYAYAFEMAPSDLLVLT 240  
 DB 181 FEELYTEAPWTETHGNLMTEROVSRMFVOCLEQSMLEIIFLYAYAFEMAPSDLLVLT 240  
 QY 241 KMFEOGFSGRQTRHLVDETMDPFVDRIGYFSALLVZGMDIESLHKCALDDBREIHF 300  
 DB 241 KMFEOGFSGRQTRHLVDETMDPFVDRIGYFSALLVZGMDIESLHKCALDDBREIHF 300  
 QY 301 AODGLICQDDCMLTFGDI PHHAPVLLAMALLRHTLNEETS SVVRKIGTAIQLVNQ 360  
 DB 301 AODGLICQDDCMLTFGDI PHHAPVLLAMALLRHTLNEETS SVVRKIGTAIQLVNQ 360  
 QY 361 YLTRLIQSLASGNDCTSTACWCYVGLSFVLTSLLEHTLIGNQODIIDTACEVLADPSL 420  
 DB 361 YLTRLIQSLASGNDCTSTACWCYVGLSFVLTSLLEHTLIGNQODIIDTACEVLADPSL 420  
 QY 421 PELFWGTEPTSGIGIILDSVCGMFPHLSPILQLRLALVSGSTAKVYSFLDKMSFYNE 480  
 DB 421 PELFWGTEPTSGIGIILDSVCGMFPHLSPILQLRLALVSGSTAKVYSFLDKMSFYNE 480  
 QY 481 LYHGKHVDVISHDGTLMRRQTPKLLYPLGQGTNLRPOGTQOVWLDDBRAYLYREYSEY 540  
 DB 481 LYHGKHVDVISHDGTLMRRQTPKLLYPLGQGTNLRPOGTQOVWLDDBRAYLYREYSEY 540  
 QY 541 SSWTLFTCEIEMLHVSTADVYIHCORVAPIIDLVHAKYSTDLSTADCLLPITSRIYML 600  
 DB 541 SSWTLFTCEIEMLHVSTADVYIHCORVAPIIDLVHAKYSTDLSTADCLLPITSRIYML 600  
 QY 601 LQRLTVISPPVDVILASCNCLITVLAARNPAKWTDLRHTGELPFVAHPVSSISQMSAS 660  
 DB 601 LQRLTVISPPVDVILASCNCLITVLAARNPAKWTDLRHTGELPFVAHPVSSISQMSAS 660  
 QY 661 GMAAGGYNLNMSSEPOGEYGVTTAFLRLITTVKGOIGSOSQGLVCVQAFVLEKMLP 720  
 DB 661 GMAAGGYNLNMSSEPOGEYGVTTAFLRLITTVKGOIGSOSQGLVCVQAFVLEKMLP 720

```
Qy 721 SYHKRVNHSRGVREGICLLIELIHAIIINLCHETDLSHSHTPSIOFLCISLAYTEAGOT 780
Db 721 SYHKRVNHSRGVREGICLLIELIHAIIINLCHETDLSHSHTPSIOFLCISLAYTEAGOT 780
Qy 781 VINIMIGVDTIDWMAAQRPSDGAEGGQGLIKIVKLAFSVTNNVIRLKPSPNVSP 840
Db 781 VINIMIGVDTIDWMAAQRPSDGAEGGQGLIKIVKLAFSVTNNVIRLKPSPNVSP 840
Qy 841 LEQALSQHGAGNNLIATLAKYIYHKDPALPRLAIOQLKLATVAPMSVYACAGNDAAA 900
Db 841 LEQALSQHGAGNNLIATLAKYIYHKDPALPRLAIOQLKLATVAPMSVYACAGNDAAA 900
Qy 901 IRDAFLRLQSKIEDMRKWMILERFLTAVETOPGELFNLLEVKQSDSKSPSLGMM 960
Db 901 IRDAFLRLQSKIEDMRKWMILERFLTAVETOPGELFNLLEVKQSDSKSPSLGMM 960
Qy 961 SCLHVAVELIDSOQDRYWCPEPLHRAAIAFLHALMODRDSAMLVATKPKFMENTSP 1020
Db 961 SCLHVAVELIDSOQDRYWCPEPLHRAAIAFLHALMODRDSAMLVATKPKFMENTSP 1020
Qy 1021 LFGTLPSPSETSEPSILETCALIMKICLEIYVVVKSGLDQSLKDTLKKFSIEKFAWMS 1080
Db 1021 LFGTLPSPSETSEPSILETCALIMKICLEIYVVVKSGLDQSLKDTLKKFSIEKFAWMS 1080
Qy 1081 GYVKSIAVHVAETGSSCTSLLEYOMLYSAMRMILITTHADIMHLDTSVVRQLFDY 1140
Db 1081 GYVKSIAVHVAETGSSCTSLLEYOMLYSAMRMILITTHADIMHLDTSVVRQLFDY 1140
Qy 1141 LDGFKALLLPVAVNCLRLGSMKCTLLILRLQWREBELGSYDEILGPTTEILEGLQADQ 1200
Db 1141 LDGFKALLLPVAVNCLRLGSMKCTLLILRLQWREBELGSYDEILGPTTEILEGLQADQ 1200
Qy 1201 QLMEXTKAKVSAFTTVLQMKEMKYSDIPOYSQVLVANCETLQOEVIALLPQTRHSLAG 1260
Db 1201 QLMEXTKAKVSAFTTVLQMKEMKYSDIPOYSQVLVANCETLQOEVIALLPQTRHSLAG 1260
Qy 1261 SATEDKQSMETDPCRSRHRDORQGVCLGLHAKELQVDEDDGSMQVTRRLPIPLTL 1320
Db 1261 SATEDKQSMETDPCRSRHRDORQGVCLGLHAKELQVDEDDGSMQVTRRLPIPLTL 1320
Qy 1321 LTTLEVSILRMKQNLHFTBATLHLTLTLARTOQATAVAGAGITOSICLPLSVYQOLSTNG 1380
Db 1321 LTTLEVSILRMKQNLHFTBATLHLTLTLARTOQATAVAGAGITOSICLPLSVYQOLSTNG 1380
Qy 1381 TAQTPSASRKSJLDAPSPGVTRLSNLSMEQLIKTLRYNPLPEALDFVGVOERTLOCLNA 1440
Db 1381 TAQTPSASRKSJLDAPSPGVTRLSNLSMEQLIKTLRYNPLPEALDFVGVOERTLOCLNA 1440
Qy 1441 VRTYQSLACEADHTVGFIIQLSNFMKEMHPLPOLMRDIOVNLGYLCOACTSLHSRK 1500
Db 1441 VRTYQSLACEADHTVGFIIQLSNFMKEMHPLPOLMRDIOVNLGYLCOACTSLHSRK 1500
Qy 1501 MLQHYLQKQNDGCLPSAQAQRVQPPSAASAAPSSSKOPADTEASEGOALHTVOYGLK 1560
Db 1501 MLQHYLQKQNDGCLPSAQAQRVQPPSAASAAPSSSKOPADTEASEGOALHTVOYGLK 1560
Qy 1561 ILSTKLALRHFTPDVQIILLDOSJDLAEYNFLFALSFTTPTPDEVAAPSGFTILLATVNV 1620
Db 1561 ILSTKLALRHFTPDVQIILLDOSJDLAEYNFLFALSFTTPTPDEVAAPSGFTILLATVNV 1620
Qy 1621 ALNMLGELDKKKEPLTOAVGLSTQABGTRTLKSLMFMENCFFYLLIQANRYLADPAVH 1680
Db 1621 ALNMLGELDKKKEPLTOAVGLSTQABGTRTLKSLMFMENCFFYLLIQANRYLADPAVH 1680
Qy 1681 PRDKORMQOELSELSTLSSLSRYFRGAPSSPATGVLPSPQGSTLSKASPPSQEPL 1740
Db 1681 PRDKORMQOELSELSTLSSLSRYFRGAPSSPATGVLPSPQGSTLSKASPPSQEPL 1740
Qy 1741 IQLVOAFVRHMOR 1753
Db 1741 IQLVOAFVRHMOR 1753
```

```
RESULT 6
US-10-719-385-13
; Sequence 13, Application US/10719385
; Publication No. US20040209284A1
; GENERAL INFORMATION:
; APPLICANT: O'Toole et al.
; TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis
; FILE REFERENCE: 22058-582
; CURRENT FILING DATE: 2003-11-21
; PRIOR FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: PCT/US03/37339
; PRIOR FILING DATE: 2002-11-21
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 1753
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-385-13

Query Match 100.0%; Score 9004; DB 17; Length 1753;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1752; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MIRSKITTSVLSFCRSSREMTIILGSAALRELSQIABELNKMWRRLLEGSLYKPPSPS 60
Db 1 MIRSKITTSVLSFCRSSREMTIILGSAALRELSQIABELNKMWRRLLEGSLYKPPSPS 60
Qy 61 SAEVYKANKVQASPLKEIGLRISKFLGIDREOSVOLQCYQOENYRGTRDSVKTVLOER 120
Db 61 SAEVYKANKVQASPLKEIGLRISKFLGIDREOSVOLQCYQOENYRGTRDSVKTVLOER 120
Qy 121 QSQALLIKIADYEEERTCILRCVLIHLTYFQDERHPYREYADCVKLEKELYSKTRQO 180
Db 121 QSQALLIKIADYEEERTCILRCVLIHLTYFQDERHPYREYADCVKLEKELYSKTRQO 180
Qy 181 FEEIYKTEAPWEHNGNLNTERQVSRNPFVQCLABQSMLEIIFLYAYFEMABSDLVLT 240
Db 181 FEEIYKTEAPWEHNGNLNTERQVSRNPFVQCLABQSMLEIIFLYAYFEMABSDLVLT 240
Qy 241 KMFKEQGFGRQTRHLYDETMDFVDRIGYFSALLIVEGMDISLHKCALDDREHLOF 300
Db 241 KMFKEQGFGRQTRHLYDETMDFVDRIGYFSALLIVEGMDISLHKCALDDREHLOF 300
Qy 301 AODGLICQDMDCMLATFGDIPHNAPVLLAMALLRHTLNPEBTSVVRKIGGTALQLNVFO 360
Db 301 AODGLICQDMDCMLATFGDIPHNAPVLLAMALLRHTLNPEBTSVVRKIGGTALQLNVFO 360
Qy 361 YLTRLQSLASGANDCTTSTRACWCYGLLSFVLTSLBLHTGNODIITDACEVLADPSL 420
Db 361 YLTRLQSLASGANDCTTSTRACWCYGLLSFVLTSLBLHTGNODIITDACEVLADPSL 420
Qy 421 PELFWGTEPSTGDIITLDSVCGMPHLLSPLOLRALVSSKSTAKKYVSLDQMSFVNE 480
Db 421 PELFWGTEPSTGDIITLDSVCGMPHLLSPLOLRALVSSKSTAKKYVSLDQMSFVNE 480
Qy 481 LYKHKPHDIVSHBEGTLMRRQTPKLLYPLAGQTMRIPOGTAVQVMDLDRAYLVWRWESY 540
Db 481 LYKHKPHDIVSHBEGTLMRRQTPKLLYPLAGQTMRIPOGTAVQVMDLDRAYLVWRWESY 540
Qy 541 SSWTLFTECEIMLHVSTADVIOHCORVPIIDLVRKISTDLSIADCLLPITSRIYML 600
Db 541 SSWTLFTECEIMLHVSTADVIOHCORVPIIDLVRKISTDLSIADCLLPITSRIYML 600
Qy 601 LQRLTVVSPVDVIVASCVNCLTVLAARNPKAWTDLRHGFLPFVAHPVSSLSQMSAB 660
Db 601 LQRLTVVSPVDVIVASCVNCLTVLAARNPKAWTDLRHGFLPFVAHPVSSLSQMSAB 660
Qy 661 GMANAGYGNLMLNSEOPEGEYGVITIAFLRLITTLVKQGLGTSQGLVPCVMFYLKEMLP 720
Db 661 GMANAGYGNLMLNSEOPEGEYGVITIAFLRLITTLVKQGLGTSQGLVPCVMFYLKEMLP 720
```

```

Db 661 GNNAGGYGNLNNSEPOGEYGVTTIAFLRLITTLVKGLGSGTOSQGLVPCVWFVLEKMLP 720
Qy 721 SYHKRNYSHGREGIOGCLILELHAIINLCHETLHSHSPSLOPLCICSIATYEAQT 780
Db 721 SYHKRNYSHGREGIOGCLILELHAIINLCHETLHSHSPSLOPLCICSIATYEAQT 780
Qy 781 VININGIGVDTIDMMAAOPRSDGAEQGGQGLIKTYKALAFSVYNNVIRLKPSPNVSP 840
Db 781 VININGIGVDTIDMMAAOPRSDGAEQGGQGLIKTYKALAFSVYNNVIRLKPSPNVSP 840
Qy 841 LEQALSOCHAGNNLIAYLAKTYHKHDPALPRLAIQILKRLATYAPMSVYACIGNDAAA 900
Db 841 LEQALSOCHAGNNLIAYLAKTYHKHDPALPRLAIQILKRLATYAPMSVYACIGNDAAA 900
Qy 901 IRDAFLTLQSKIEEMRIKWMLEFLTYAVETOPGLIELFNLKXKSDSKSPSGMW 960
Db 901 IRDAFLTLQSKIEEMRIKWMLEFLTYAVETOPGLIELFNLKXKSDSKSPSGMW 960
Qy 961 SCLHNVLELIDSOQDRYWCPEPLHRAAIAFLHALMODRDSAMLVLTKEKFEWENLTSP 1020
Db 961 SCLHNVLELIDSOQDRYWCPEPLHRAAIAFLHALMODRDSAMLVLTKEKFEWENLTSP 1020
Qy 1021 LFGTISPSSESEPSILFETCALIMKICLETIYVYVKGSLDGLKDTLKKSEIERKPAWS 1080
Db 1021 LFGTISPSSESEPSILFETCALIMKICLETIYVYVKGSLDGLKDTLKKSEIERKPAWS 1080
Qy 1081 GYVKSIAVHAETESSCTSLLEYOMLVSAWMLIITTHADIMHLDTSVVRQLFLDV 1140
Db 1081 GYVKSIAVHAETESSCTSLLEYOMLVSAWMLIITTHADIMHLDTSVVRQLFLDV 1140
Qy 1141 LDGTALLLVPAVNCRLGSMKCTLLIILRQWRELGSDVEILGPTEILEGVLQADQ 1200
Db 1141 LDGTALLLVPAVNCRLGSMKCTLLIILRQWRELGSDVEILGPTEILEGVLQADQ 1200
Qy 1201 QLMETKAKVSAFTTVQMKMKVSDIPOYSQVLAVNCETLQEBVALFQGTSHSLAG 1260
Db 1201 QLMETKAKVSAFTTVQMKMKVSDIPOYSQVLAVNCETLQEBVALFQGTSHSLAG 1260
Qy 1261 SATEDKDSMETDSCSRSHRDQDVCVLGLAKELCEVDEDDGSMQVTRRLPILPTL 1320
Db 1261 SATEDKDSMETDSCSRSHRDQDVCVLGLAKELCEVDEDDGSMQVTRRLPILPTL 1320
Qy 1321 LTTLEVSIRMKONLHFTBATLHLITLARTQOGATVAVAGAGITOSICPLISVYQSTNG 1380
Db 1321 LTTLEVSIRMKONLHFTBATLHLITLARTQOGATVAVAGAGITOSICPLISVYQSTNG 1380
Qy 1381 TAOTPSASRKSILDAPSWPGVYRLSWSLMEOLLKTLRYNPLPEALDPYGVHOERTLQCLNA 1440
Db 1381 TAOTPSASRKSILDAPSWPGVYRLSWSLMEOLLKTLRYNPLPEALDPYGVHOERTLQCLNA 1440
Qy 1441 VRTVQSIACLEBADHTVGFILQLSNFMKEWPHLPOLMRDIQVNLGYLQCACTSLHSRK 1500
Db 1441 VRTVQSIACLEBADHTVGFILQLSNFMKEWPHLPOLMRDIQVNLGYLQCACTSLHSRK 1500
Qy 1501 MLQOYTLQKNDGIPSAVAQVQRPSSAASAPSSSKOPADTASRQALHTYQGLK 1560
Db 1501 MLQOYTLQKNDGIPSAVAQVQRPSSAASAPSSSKOPADTASRQALHTYQGLK 1560
Qy 1561 ILSTKLALRHFTPDVCOILLDQSLDLAEYNFALSTFTTPDSEVAPSGTLLATVNV 1620
Db 1561 ILSTKLALRHFTPDVCOILLDQSLDLAEYNFALSTFTTPDSEVAPSGTLLATVNV 1620
Qy 1621 ALNMLGELDKKKEPLTOAVGLSTQAEGRITLKSILMTMENCYLLISQAMRYLRDPAVH 1680
Db 1621 ALNMLGELDKKKEPLTOAVGLSTQAEGRITLKSILMTMENCYLLISQAMRYLRDPAVH 1680
Qy 1681 PRDQRMKQELSSSTLSSISRYFRGASSPATVLPBPQKSTSLASAPESQEP 1740
Db 1681 PRDQRMKQELSSSTLSSISRYFRGASSPATVLPBPQKSTSLASAPESQEP 1740
Qy 1741 IQLVQAFVRHMQR 1753
Db 1741 IQLVQAFVRHMQR 1753

```

```

RESULT 7
US-10-719-385-14
; Sequence 14, Application US/10719385
; Publication No. US20040209284A1
; GENERAL INFORMATION:
; APPLICANT: O'Toole et al.
; TITLE OR INVENTION: Composition and Method for Treating Lupus Nephritis
; FILE REFERENCE: 22058-582
; CURRENT APPLICATION NUMBER: US/10/719,385
; PRIOR FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: PCT/US03/37339
; PRIOR FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: 60/428,094
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; LENGTH: 1753
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-385-14

```

Query Match 100.0%; Score 9004; DB 17; Length 1753;

Best Local Similarity 99.9%; Pred. No. 0; Matches 1752; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

Qy 1 MIRSKTSTVLSFCRSREIMTLLGRSALRELSQIEAEINKMRRLLEGLSTYKPPSPS 60
Db 1 MIRSKTSTVLSFCRSREIMTLLGRSALRELSQIEAEINKMRRLLEGLSTYKPPSPS 60
Qy 61 SAEKVRANKVAPLKLGLRISKPLGLDEQSVOLQCYLQDYGTRPSVTYQDER 120
Db 61 SAEKVRANKVAPLKLGLRISKPLGLDEQSVOLQCYLQDYGTRPSVTYQDER 120
Qy 121 QSQALIKIADYYEERTCIARCVLHLTYFQDERHPYREYADCVDKLEKELVSKYRQ 180
Db 121 QSQALIKIADYYEERTCIARCVLHLTYFQDERHPYREYADCVDKLEKELVSKYRQ 180
Qy 181 FEEIYKTEATYTWTHGNLWTEROVSRWFOCLREQSMLEIFLTYAYFEMASDILVLT 240
Db 181 FEEIYKTEATYTWTHGNLWTEROVSRWFOCLREQSMLEIFLTYAYFEMASDILVLT 240
Qy 241 KMFKEQGFSGRNRLHVDETMDPFVDRIGYFSALILVEGMDIESLHKCLDNRRELHQP 300
Db 241 KMFKEQGFSGRNRLHVDETMDPFVDRIGYFSALILVEGMDIESLHKCLDNRRELHQP 300
Qy 301 AODGLICQMDICMLTFGDI PHHAPVLLAWALLRHTLNPEETSSVVRKIGGTAIQNLVFP 360
Db 301 AODGLICQMDICMLTFGDI PHHAPVLLAWALLRHTLNPEETSSVVRKIGGTAIQNLVFP 360
Qy 361 YLRLRLQSLASGNDCTTSTACMCVYGLSFVLTSELTGNGOODITPACVLDPSL 420
Db 361 YLRLRLQSLASGNDCTTSTACMCVYGLSFVLTSELTGNGOODITPACVLDPSL 420
Qy 421 PELFWGTEPTSGILILDSVCGMFPHLSPILQLRLALVSGKSTAKKYSFLDKMSFYNE 480
Db 421 PELFWGTEPTSGILILDSVCGMFPHLSPILQLRLALVSGKSTAKKYSFLDKMSFYNE 480
Qy 481 LYHGHKPDVYSHEDGTLMRQTPKLLYPIGGQTNLRIPOGTGQVMLDRAVYVRWEYSY 540
Db 481 LYHGHKPDVYSHEDGTLMRQTPKLLYPIGGQTNLRIPOGTGQVMLDRAVYVRWEYSY 540
Qy 541 SSTWLTFCREMLHNVSTADVYOHQORVKPIIDLVHXYISDLSIADCLPITSRYNL 600
Db 541 SSTWLTFCREMLHNVSTADVYOHQORVKPIIDLVHXYISDLSIADCLPITSRYNL 600
Qy 601 LQRLTVISPPVDVIASCNCLTVLAARNPAKWTDLRHTGFLPFVAHPVSSISQMSISAB 660
Db 601 LQRLTVISPPVDVIASCNCLTVLAARNPAKWTDLRHTGFLPFVAHPVSSISQMSISAB 660
Qy 661 GNNAGGYGNLNNSEPOGEYGVTTIAFLRLITTLVKGLGSGTOSQGLVPCVWFVLEKMLP 720

```

```

Db 661 GMAAGYGNLMSNGPQGEYGVITIAFLRLITTLVKGGLSGTQSGVLPCWFMVLKXLP 720
Qy 721 SYHKRRVNSHVRBOIGCLILIELHAIINLCHETDLSHSHPSLOFLCISLAYEAGOT 780
Db 721 SYHKRRVNSHVRBOIGCLILIELHAIINLCHETDLSHSHPSLOFLCISLAYEAGOT 780
Qy 781 VINIMGIGVDITIDWMAOPRSDEAGGOGOLIKTKYKLA.PSVYNNVYRLKPPENVSP 840
Db 781 VINIMGIGVDITIDWMAOPRSDEAGGOGOLIKTKYKLA.PSVYNNVYRLKPPENVSP 840
Qy 841 LEQALSOHGAHGNMLIAVLAKYIYKHDPALPRLAIQULKELATVAPNSVYACLGNDAAA 900
Db 841 LEQALSOHGAHGNMLIAVLAKYIYKHDPALPRLAIQULKELATVAPNSVYACLGNDAAA 900
Qy 901 IRDAFLTRLOSKIEEMRIKMWLEFLTYAVETOPGLIFLNLBYKQSDSKESPLGMW 960
Db 901 IRDAFLTRLOSKIEEMRIKMWLEFLTYAVETOPGLIFLNLBYKQSDSKESPLGMW 960
Qy 961 SCLHAYVELIDSQOQDRYWCPEPLHRAAIAFLHAIAMODRPSAMLVLRKPKPMENLTS 1020
Db 961 SCLHAYVELIDSQOQDRYWCPEPLHRAAIAFLHAIAMODRPSAMLVLRKPKPMENLTS 1020
Qy 1021 LFGTLPSPSETSEPSILETCALIMKICLEIYVVVKSJLDSQKXDTLKKFSIEKRFAYMS 1080
Db 1021 LFGTLPSPSETSEPSILETCALIMKICLEIYVVVKSJLDSQKXDTLKKFSIEKRFAYMS 1080
Qy 1081 GYVKSALAVHVAETBESSCTSLLEYOMLVANRMLIITTTADIMHLDTSVVRQLPVDY 1140
Db 1081 GYVKSALAVHVAETBESSCTSLLEYOMLVANRMLIITTTADIMHLDTSVVRQLPVDY 1140
Qy 1141 LDGTRALLVPAVNCRLGSMKCTLLILROMREIGSVDEILGPTLLEGLQADQ 1200
Db 1141 LDGTRALLVPAVNCRLGSMKCTLLILROMREIGSVDEILGPTLLEGLQADQ 1200
Qy 1201 QLMEXTKAKVSAFITYLQMKEMKVS.DI.PQYSQVLVANCETLQEBEVALPQTRSHSLAG 1260
Db 1201 QLMEXTKAKVSAFITYLQMKEMKVS.DI.PQYSQVLVANCETLQEBEVALPQTRSHSLAG 1260
Qy 1261 SATEDKDSMETDDCSRSRHRDQDVCYGLHAKELCEVDEDDGSMQVTRRLPILPTL 1320
Db 1261 SATEDKDSMETDDCSRSRHRDQDVCYGLHAKELCEVDEDDGSMQVTRRLPILPTL 1320
Qy 1321 LTTLEVSIRMKONHAFTEATLHLLTLARTOOGAFAVAGAGITOSICPLSVYQLSTNG 1380
Db 1321 LTTLEVSIRMKONHAFTEATLHLLTLARTOOGAFAVAGAGITOSICPLSVYQLSTNG 1380
Qy 1381 TAQTPSASRKS.LDAPSWPGVYRLSWSLMEQLKTLRYNFPLEALDPFVGHQERTLQCLNA 1440
Db 1381 TAQTPSASRKS.LDAPSWPGVYRLSWSLMEQLKTLRYNFPLEALDPFVGHQERTLQCLNA 1440
Qy 1441 VRTVOSLACLEBADHTVGFIQLSNFMKEWHFHLPOLMRDIOVNIQVYCOACTSLSHSRK 1500
Db 1441 VRTVOSLACLEBADHTVGFIQLSNFMKEWHFHLPOLMRDIOVNIQVYCOACTSLSHSRK 1500
Qy 1501 MLQHYLQKNGDGLPSAFAORVORPPSAASAAPSSKOPADTEASEOALHTVYGLK 1560
Db 1501 MLQHYLQKNGDGLPSAFAORVORPPSAASAAPSSKOPADTEASEOALHTVYGLK 1560
Qy 1561 ILSTKLALRHFTPDVCOILLDOSIDLAEYNFLPALSTTPTFDEEVA.PSGFTLLATVNV 1620
Db 1561 ILSTKLALRHFTPDVCOILLDOSIDLAEYNFLPALSTTPTFDEEVA.PSGFTLLATVNV 1620
Qy 1621 ALNMLGEIDKKEP.LTOAVGLSTQAEGRITLKS.LMFMENC.FYLLIQAMRYLDPVH 1680
Db 1621 ALNMLGEIDKKEP.LTOAVGLSTQAEGRITLKS.LMFMENC.FYLLIQAMRYLDPVH 1680
Qy 1681 PRDKORMKOELSSSLSTLSSLSRYFRGAPSPATGVLPSPQGSTLSKASPSQSP.L 1740
Db 1681 PRDKORMKOELSSSLSTLSSLSRYFRGAPSPATGVLPSPQGSTLSKASPSQSP.L 1740
Qy 1741 IOLVOAFVRHMQR 1753

```

```

Db 1741 IOLVOAFVRHMQR 1753
RESULT 8
US-10-719-385-15
; Sequence 15, Application US/10719385
; Publication No. US20040209284A1
; GENERAL INFORMATION:
; APPLICANT: O'Toole et al.
; TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis
; FILE REFERENCE: 22058-582
; CURRENT FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: PCT/US03/37339
; PRIOR FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: 60/428,094
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 15
; LENGTH: 1753
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-385-15
Query Match 100.0%; Score 9003; DB 17; Length 1753;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1751; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MIRSKITSVLSFCRSSRHELTILLGRSALRELQIEALNKHWRLLGLSYKPPSPS 60
Db 1 MIRSKITSVLSFCRSSRHELTILLGRSALRELQIEALNKHWRLLGLSYKPPSPS 60
Qy 61 SAEYKANKDVASPLKEGLRISKFLGIDEQSVOQLQCTYQEBYRGTRDSVKTVLQDER 120
Db 61 SAEYKANKDVASPLKEGLRISKFLGIDEQSVOQLQCTYQEBYRGTRDSVKTVLQDER 120
Qy 121 OSQALIKIADYVYEBRCLRCVLAHLTYRQDRHPRVAYACVDELKELVSKYRQ 180
Db 121 OSQALIKIADYVYEBRCLRCVLAHLTYRQDRHPRVAYACVDELKELVSKYRQ 180
Qy 181 FEELYKTEAPTWETHGNLMTEROVS.RWFVQCLAREOSMLEIIFLYAYAFENAPSDLVLT 240
Db 181 FEELYKTEAPTWETHGNLMTEROVS.RWFVQCLAREOSMLEIIFLYAYAFENAPSDLVLT 240
Qy 241 KMFEKQGFSGQTRNHLVDETMDFVDRIGYFSALILVEGMDISLHKCALDDRRRLHOF 300
Db 241 KMFEKQGFSGQTRNHLVDETMDFVDRIGYFSALILVEGMDISLHKCALDDRRRLHOF 300
Qy 301 AODGLICQDMQCLMLTGGDI.PHNA.PVLLAWAL.LHHTLNPEETS.VVRKIGGTAIQLVNFQ 360
Db 301 AODGLICQDMQCLMLTGGDI.PHNA.PVLLAWAL.LHHTLNPEETS.VVRKIGGTAIQLVNFQ 360
Qy 361 YLFRLLQSLASGANDCTTSTACMCVYGLSLVLTSLBLHTLGNQODIIDTA.CEVLADPSL 420
Db 361 YLFRLLQSLASGANDCTTSTACMCVYGLSLVLTSLBLHTLGNQODIIDTA.CEVLADPSL 420
Qy 421 PELFWGTEPTSGGLIILDSVGMPPHLSPLQLRLALVSKSTAKKYSLDKKSFYNE 480
Db 421 PELFWGTEPTSGGLIILDSVGMPPHLSPLQLRLALVSKSTAKKYSLDKKSFYNE 480
Qy 481 LYKHKPHDVISHEDGTLMRQTPKLLVPLGGQTMRLIPQGVGVOMLDDRA.YLVWMEYSY 540
Db 481 LYKHKPHDVISHEDGTLMRQTPKLLVPLGGQTMRLIPQGVGVOMLDDRA.YLVWMEYSY 540
Qy 541 SSWTLFTCEIMELHVVSTADVIOHQKRVKPIDLVHKNVISTDLSIADCLLPITSRIYML 600
Db 541 SSWTLFTCEIMELHVVSTADVIOHQKRVKPIDLVHKNVISTDLSIADCLLPITSRIYML 600
Qy 601 LQRLITTVISPPVDVIASGVNCLTYLAARNPAKWMTDLRHGTF.LPVAHPVSSLSQMSAE 660
Db 601 LQRLITTVISPPVDVIASGVNCLTYLAARNPAKWMTDLRHGTF.LPVAHPVSSLSQMSAE 660

```



QY 661 GNNAGYGNLLMNSPOGEVGTIAFLRLITTLVKGOLSTQSGVLPCVWFVLEKMLP 720  
 DB 661 GNNAGYGNLLMNSPOGEVGTIAFLRLITTLVKGOLSTQSGVLPCVWFVLEKMLP 720  
 QY 721 SYHKRYNSHGVREIOGCLILELHAILNLCHETDLSSHTPSLOFLCISIAYTEAGOT 780  
 DB 721 SYHKRYNSHGVREIOGCLILELHAILNLCHETDLSSHTPSLOFLCISIAYTEAGOT 780  
 QY 781 VININGIGVDTIDMMAAOPRSDGABGGQGGOLIKTYKLAFTYNNNTYRLKPPENYSP 840  
 DB 781 VININGIGVDTIDMMAAOPRSDGABGGQGGOLIKTYKLAFTYNNNTYRLKPPENYSP 840  
 QY 841 LEQALSOHAGHNNLIAYLAKYIYKHDPALPRLAIQILKRLATVAPSVYACIGNDAAA 900  
 DB 841 LEQALSOHAGHNNLIAYLAKYIYKHDPALPRLAIQILKRLATVAPSVYACIGNDAAA 900  
 QY 901 IRDALVTLRQSKIEEMRIKMWLEFLTVAVETQPGILFLNLEYKDSGDSKESFLGMV 960  
 DB 901 IRDALVTLRQSKIEEMRIKMWLEFLTVAVETQPGILFLNLEYKDSGDSKESFLGMV 960  
 QY 961 SCLHAYLELIDBQOQDRYWCPELHRAAIAFLHLMODRDSAMVLTTRKPFMENLTSP 1020  
 DB 961 SCLHAYLELIDBQOQDRYWCPELHRAAIAFLHLMODRDSAMVLTTRKPFMENLTSP 1020  
 QY 1021 LFGTLPSPSETSEPSILETCALIMKICLEIYVVVKGSLDQSLKDTLKFSIEKRFAYMS 1080  
 DB 1021 LFGTLPSPSETSEPSILETCALIMKICLEIYVVVKGSLDQSLKDTLKFSIEKRFAYMS 1080  
 QY 1081 GYVSLAHVAETESSCTSLIEYOMLVSAWPMILITTHADIMHLDTSVVRQOLFV 1140  
 DB 1081 GYVSLAHVAETESSCTSLIEYOMLVSAWPMILITTHADIMHLDTSVVRQOLFV 1140  
 QY 1141 LDGTALLLVPAVNCRLGSMKCTLLILNOMREIGSVDEILGPTLEILEGVLQADQ 1200  
 DB 1141 LDGTALLLVPAVNCRLGSMKCTLLILNOMREIGSVDEILGPTLEILEGVLQADQ 1200  
 QY 1201 QLMETKAKVSAFTTVQMKEMKSDIPQYSQVLANCETLQEBVIALPQTRHSIALG 1260  
 DB 1201 QLMETKAKVSAFTTVQMKEMKSDIPQYSQVLANCETLQEBVIALPQTRHSIALG 1260  
 QY 1261 SATEDKDSMETDSCRSRHRDORQDVCYLGHAKELCEVDEBDSMLOVTRRLPIPL 1320  
 DB 1261 SATEDKDSMETDSCRSRHRDORQDVCYLGHAKELCEVDEBDSMLOVTRRLPIPL 1320  
 QY 1321 LTTLEVSIRMFONLHFTBATLHLITLARTQOGATAVAGAGITQSCILPLSVYQLSTNG 1380  
 DB 1321 LTTLEVSIRMFONLHFTBATLHLITLARTQOGATAVAGAGITQSCILPLSVYQLSTNG 1380  
 QY 1381 TAOTPSASRSKSLDAPSWGVYRLSMSLMEQLKTLRYNFLPEALDPGVCHOERTLOCLNA 1440  
 DB 1381 TAOTPSASRSKSLDAPSWGVYRLSMSLMEQLKTLRYNFLPEALDPGVCHOERTLOCLNA 1440  
 QY 1441 VRTQVSLACLEBAHTVGFILQLSNFMKEWHFHLPOLMRDIQVNLGYLCOACTSLHSRK 1500  
 DB 1441 VRTQVSLACLEBAHTVGFILQLSNFMKEWHFHLPOLMRDIQVNLGYLCOACTSLHSRK 1500  
 QY 1501 MLOHYLQKNGDGLPSAVAQEVQRPSSASAAPSSSQPADTASQOALHTVOYGLK 1560  
 DB 1501 MLOHYLQKNGDGLPSAVAQEVQRPSSASAAPSSSQPADTASQOALHTVOYGLK 1560  
 QY 1561 ILSTKLALHRTFDVCOIILLDOSLDLAEVNFALSTFTTPRSEVAPSGTILAVNV 1620  
 DB 1561 ILSTKLALHRTFDVCOIILLDOSLDLAEVNFALSTFTTPRSEVAPSGTILAVNV 1620  
 QY 1621 ALNMLGELDKKKEPLTQAVGSTQABGRITLKSILMTMENCFLILSQAMRYLRDPAVH 1680  
 DB 1621 ALNMLGELDKKKEPLTQAVGSTQABGRITLKSILMTMENCFLILSQAMRYLRDPAVH 1680  
 QY 1681 PRDQKMKOELSSSLSTLSLSYFRGAPSSPATGVLPSPQCKSTLSKASPSQSEPL 1740  
 DB 1681 PRDQKMKOELSSSLSTLSLSYFRGAPSSPATGVLPSPQCKSTLSKASPSQSEPL 1740  
 QY 1741 IQLVQAFVRHMQR 1753

DB 1741 IQLVQAFVRHMQR 1753  
 RESULT 9  
 US-10-719-385-6  
 ; Sequence 6, Application US/10719385  
 ; Publication No. US20040209284A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: O'toole et al.  
 ; TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis  
 ; FILE REFERENCE: 22058-582  
 ; CURRENT APPLICATION NUMBER: US/10/719,385  
 ; CURRENT FILING DATE: 2003-11-21  
 ; PRIOR APPLICATION NUMBER: PCT/US03/37339  
 ; PRIOR FILING DATE: 2003-11-21  
 ; PRIOR APPLICATION NUMBER: 60/428,094  
 ; PRIOR FILING DATE: 2002-11-21  
 ; NUMBER OF SEQ ID NOS: 26  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 6  
 ; LENGTH: 1753  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-10-719-385-6

Query Match 99.9%; Score 9002; DB 17; Length 1753;  
 Best Local Similarity 99.9%; Pred. No. 0;  
 Matches 1753; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MIRSKITSVLSFCRSSREBLMTILLGRSALRELSQIEAEINKMWRLLBGLSYKKPPSPS 60  
 DB 1 MIRSKITSVLSFCRSSREBLMTILLGRSALRELSQIEAEINKMWRLLBGLSYKKPPSPS 60  
 QY 61 SAEKVANKVAPLKLRLGRLSKFJGLDBEQQVOLLOCYLQBDYGRTRDSVKTVALDER 120  
 DB 61 SAEKVANKVAPLKLRLGRLSKFJGLDBEQQVOLLOCYLQBDYGRTRDSVKTVALDER 120  
 QY 121 QSOALILKIDVYYEERTCLRCVHLHLYFODEBRYPREVADCVDKELVSKYRQO 180  
 DB 121 QSOALILKIDVYYEERTCLRCVHLHLYFODEBRYPREVADCVDKELVSKYRQO 180  
 QY 181 FEEIYKTEAPWETHGNLMTERQVSRWFVQCLREQSMLLEIFLYYAYEMABSDLLVLT 240  
 DB 181 FEEIYKTEAPWETHGNLMTERQVSRWFVQCLREQSMLLEIFLYYAYEMABSDLLVLT 240  
 QY 241 KMFKEQFGSRQNRHLVDETMDFVDRIGYPSALLVGEQMDIESLHKCALDRRELHOF 300  
 DB 241 KMFKEQFGSRQNRHLVDETMDFVDRIGYPSALLVGEQMDIESLHKCALDRRELHOF 300  
 QY 301 AODGLICODMDCMLTFGDI PHHA PVLAMALLRHTLNPEETSSVVRKIGGTAIQNLVFO 360  
 DB 301 AODGLICODMDCMLTFGDI PHHA PVLAMALLRHTLNPEETSSVVRKIGGTAIQNLVFO 360  
 QY 361 YLTRLQSLASGGNDCTTSTACMCVYGLISFVLTSLHTLGNQODI IDTACEVLADPSL 420  
 DB 361 YLTRLQSLASGGNDCTTSTACMCVYGLISFVLTSLHTLGNQODI IDTACEVLADPSL 420  
 QY 421 PELFMGTPEPSGILILDSVCGMFPRLSPILQILBALVSGSTAKKVSFLDKMSFYNE 480  
 DB 421 PELFMGTPEPSGILILDSVCGMFPRLSPILQILBALVSGSTAKKVSFLDKMSFYNE 480  
 QY 481 LYHGRKHDAVISHEDGTLMRRQTPKLLYPLGQOTNLRI PGTVQGVMLDRAVLYVREYSY 540  
 DB 481 LYHGRKHDAVISHEDGTLMRRQTPKLLYPLGQOTNLRI PGTVQGVMLDRAVLYVREYSY 540  
 QY 541 SSWTLFTCEIEMLHYVSTADYI OHCQRYKPIIDLVHKYISTDLSTADCLLPITSRIVML 600  
 DB 541 SSWTLFTCEIEMLHYVSTADYI OHCQRYKPIIDLVHKYISTDLSTADCLLPITSRIVML 600  
 QY 601 LORLTVIAPPVAVLASCNCLTVLAARNPAKVTMDLRTGFLPFAHVVSSLSQWISAE 660  
 DB 601 LORLTVIAPPVAVLASCNCLTVLAARNPAKVTMDLRTGFLPFAHVVSSLSQWISAE 660



QY 661 GNNAGVGNLNMSEPOGEGVTTAFLRLITTLVKGQSGTOSGGLVPCWFLKEMLP 720  
DB 661 GNNAGVGNLNMSEPOGEGVTTAFLRLITTLVKGQSGTOSGGLVPCWFLKEMLP 720  
QY 721 SYHKRYNSHGVRSQIGCLIELIHAITNLCHETDHSHPSPLOFLCTCSLATYAGQT 780  
DB 721 SYHKRYNSHGVRSQIGCLIELIHAITNLCHETDHSHPSPLOFLCTCSLATYAGQT 780  
QY 781 VININGIGVDTIDWMAAOPRSDGEGGQGLIKTVKLAFSVTNNVIRLKPSPNVSP 840  
DB 781 VININGIGVDTIDWMAAOPRSDGEGGQGLIKTVKLAFSVTNNVIRLKPSPNVSP 840  
QY 841 LEQALSHGAGNNLIATLAKTYHKHDPALPRLAIQILKRLATYAPSPVACGNDAAA 900  
DB 841 LEQALSHGAGNNLIATLAKTYHKHDPALPRLAIQILKRLATYAPSPVACGNDAAA 900  
QY 901 IRDAFLTRLOSKIEDMRKMWLEFLVAVEPOGLIELFLNLEVKDSDGSKESFLGMM 960  
DB 901 IRDAFLTRLOSKIEDMRKMWLEFLVAVEPOGLIELFLNLEVKDSDGSKESFLGMM 960  
QY 961 SCLHAYLELIDSOQODRYWCPPLHRAAIAFLHLMODRDSAMLVLRKPFMENTTSP 1020  
DB 961 SCLHAYLELIDSOQODRYWCPPLHRAAIAFLHLMODRDSAMLVLRKPFMENTTSP 1020  
QY 1021 LFGTLSPSETSEBSILETCALIMKIICLETIYVYVKGSLDSDLKDTLKKESIEKRFAPWS 1080  
DB 1021 LFGTLSPSETSEBSILETCALIMKIICLETIYVYVKGSLDSDLKDTLKKESIEKRFAPWS 1080  
QY 1081 GYVSLAVHVAETGSSCTSLLEYOMLVSAWRMLLIITTHADIMHLDTSVVRQOLFVDV 1140  
DB 1081 GYVSLAVHVAETGSSCTSLLEYOMLVSAWRMLLIITTHADIMHLDTSVVRQOLFVDV 1140  
QY 1141 LDGTRALLLVASVNCRLGSMKCTLLIILRQWRBELGSVDEILGPTLEILGVLQADQ 1200  
DB 1141 LDGTRALLLVASVNCRLGSMKCTLLIILRQWRBELGSVDEILGPTLEILGVLQADQ 1200  
QY 1201 QLMERTKAKVSAFTIYVQMKEMKVSDFPQSVQVLANCETLOEBEVIALPQOTRHSALG 1260  
DB 1201 QLMERTKAKVSAFTIYVQMKEMKVSDFPQSVQVLANCETLOEBEVIALPQOTRHSALG 1260  
QY 1261 SATEDKDSMETDSCSRSHRDQDQVCYLGHAKELCEVEDSDSMQVTRRLPIEPTL 1320  
DB 1261 SATEDKDSMETDSCSRSHRDQDQVCYLGHAKELCEVEDSDSMQVTRRLPIEPTL 1320  
QY 1321 LTTLEVSILRMKQNLHFTBATLHLITLARTOAGATAVAGAGITOSICPLLSVYQLSTNG 1380  
DB 1321 LTTLEVSILRMKQNLHFTBATLHLITLARTOAGATAVAGAGITOSICPLLSVYQLSTNG 1380  
QY 1381 TAQTPSASRSKSLDAPSWPGVYRLSWSLMEQLKTRYNFLPEALDFVGHBERLQOCINA 1440  
DB 1381 TAQTPSASRSKSLDAPSWPGVYRLSWSLMEQLKTRYNFLPEALDFVGHBERLQOCINA 1440  
QY 1441 VRTVQSLACEADHTVGFIIQLSNFMKEMPHLPQLMRDIOVNLGVYCOACTSLHSRK 1500  
DB 1441 VRTVQSLACEADHTVGFIIQLSNFMKEMPHLPQLMRDIOVNLGVYCOACTSLHSRK 1500  
QY 1501 MLQHYLQNKNGDGLPSAQAORVORPPSAASAAPSSSKOPADTEASEQOALHTVQYGLK 1560  
DB 1501 MLQHYLQNKNGDGLPSAQAORVORPPSAASAAPSSSKOPADTEASEQOALHTVQYGLK 1560  
QY 1561 ILSTKLALRHFTPDVCOILLDOSJDLAEYNFLPALSTFTPTFSEVAAPSGTLLATYAV 1620  
DB 1561 ILSTKLALRHFTPDVCOILLDOSJDLAEYNFLPALSTFTPTFSEVAAPSGTLLATYAV 1620  
QY 1621 ALNMLGELDKKKEPLTOAVGISTOAGRTTLKSLMFMEKCFYLLISQARVYLRDPAVH 1680  
DB 1621 ALNMLGELDKKKEPLTOAVGISTOAGRTTLKSLMFMEKCFYLLISQARVYLRDPAVH 1680  
QY 1681 PRDKORMKQBLSELSTLSSLSRYFRGAPSPATGVLPSPQKSTLSKASPESSQBL 1740  
DB 1681 PRDKORMKQBLSELSTLSSLSRYFRGAPSPATGVLPSPQKSTLSKASPESSQBL 1740

QY 1741 IQLVQAFVRHMOR 1753  
DB 1741 IQLVQAFVRHMOR 1753  
RESULT 10  
US-10-719-385-10  
; Sequence 10, Application US/10719385  
; Publication No. US20040209284A1  
; GENERAL INFORMATION:  
; APPLICANT: O'Toole et al.  
; TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis  
; FILE REFERENCE: 22058-582  
; CURRENT APPLICATION NUMBER: US/10/719,385  
; PRIOR FILING DATE: 2003-11-21  
; PRIOR APPLICATION NUMBER: PCT/US03/37339  
; PRIOR FILING DATE: 2002-11-21  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 10  
; LENGTH: 1753  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-719-385-10  
Query Match 99.9%; Score 9002; DB 17; Length 1753;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1752; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MIRSKITSVLSFRRSSRELMTILLGRSALRELSQIEALNKMRRLLLEGISYKPPSPS 60  
DB 1 MIRSKITSVLSFRRSSRELMTILLGRSALRELSQIEALNKMRRLLLEGISYKPPSPS 60  
QY 61 SAEKYNKANKVADSLKEJGLRISKFLGDEBQSVOLLOCYQOEDYRGTRDSVKTVLQDER 120  
DB 61 SAEKYNKANKVADSLKEJGLRISKFLGDEBQSVOLLOCYQOEDYRGTRDSVKTVLQDER 120  
QY 121 QSQALLIKIADYYEERTCIIRCVLHLITFYQDERHPYRVEYADCVDLKEKELVSKTRQ 180  
DB 121 QSQALLIKIADYYEERTCIIRCVLHLITFYQDERHPYRVEYADCVDLKEKELVSKTRQ 180  
QY 181 FEEIYKTEAPWETHGNLMTEROYSRWFVQCLREOSMLLEIFLYVAIFENAPSDLVLT 240  
DB 181 FEEIYKTEAPWETHGNLMTEROYSRWFVQCLREOSMLLEIFLYVAIFENAPSDLVLT 240  
QY 241 KMFKEQGSRSQTRHLVDETMDFVDRIGYFSALLIVEGMDISLHKCALDDRREHLQF 300  
DB 241 KMFKEQGSRSQTRHLVDETMDFVDRIGYFSALLIVEGMDISLHKCALDDRREHLQF 300  
QY 301 AODGELICODMDCMLTREGDIPHNAVLLAMALLRHTINPEETSSVVRKIGSTAIQLNVQ 360  
DB 301 AODGELICODMDCMLTREGDIPHNAVLLAMALLRHTINPEETSSVVRKIGSTAIQLNVQ 360  
QY 361 YLTRLOSLASGNDCTTSTACVCYGLISVLTSLBHTIGNOODITDTACEVLADPSL 420  
DB 361 YLTRLOSLASGNDCTTSTACVCYGLISVLTSLBHTIGNOODITDTACEVLADPSL 420  
QY 421 PELFMGTEPTSGILIIIDSVCGMPHLLSPILQILRALVSGKSTAKKVSFLDKMSFYNE 480  
DB 421 PELFMGTEPTSGILIIIDSVCGMPHLLSPILQILRALVSGKSTAKKVSFLDKMSFYNE 480  
QY 481 LYKHKPHDVISHEDGTLMRQTPRLVPLGGQTNLRIPQGTVGQVMLDDRAYLVREWSY 540  
DB 481 LYKHKPHDVISHEDGTLMRQTPRLVPLGGQTNLRIPQGTVGQVMLDDRAYLVREWSY 540  
QY 541 SSWTLFCEIEMLLHVVSTADVIOHCORVXPIIDLVHKVISTDLSIADCLLPITSRIYML 600  
DB 541 SSWTLFCEIEMLLHVVSTADVIOHCORVXPIIDLVHKVISTDLSIADCLLPITSRIYML 600  
QY 601 LQRTTVIAPSVVDIASCVNCLTVLAARNPAKVTMDLRHTGFLPFVAHPVSSLQMTISAE 660  
DB 601 LQRTTVIAPSVVDIASCVNCLTVLAARNPAKVTMDLRHTGFLPFVAHPVSSLQMTISAE 660

[illegible]

```

OY      1741  IOLVOAFVRFHMOR 1753
        |||||||
DB      1741  IOLVOAFVRFHMOR 1753

RESULT 11
        US-10-719-385-16
        ? Sequence 16, Application US/10719385
        ? Publication No. US200402093284A1
        ? GENERAL INFORMATION:
        ? APPLICANT: O'Toole et al.
        ? TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis
        ? FILE REFERENCE: 22058-582
        ? CURRENT APPLICATION NUMBER: US/10/719.385
        ? CURRENT FILING DATE: 2003-11-21
        ? PRIOR APPLICATION NUMBER: PCT/US03/37339
        ? PRIOR FILING DATE: 2003-11-21
        ? PRIOR APPLICATION NUMBER: 60/428,094
        ? PRIOR FILING DATE: 2002-11-21
        ? NUMBER OF SEQ ID NOS: 26
        ? SOFTWARE: PatentIn Ver. 2.1
        ? SEQ ID NO: 16
        ? LENGTH: 1753
        ? TYPE: PRT
        ? ORGANISM: Homo sapiens
        ? US-10-719-385-16

```

Query Match	Best Local Similarity	99.9%	Score 9001;	DB 17;	Length 1753;			
Matches 1751;	Conservative	2;	Mismatches	0;	Indels	0;	Gaps	0;
QY	1	MIRKSKITSVLSFCSRREIMWTILGRSALRELSQIAEALNKMRRLLEGSIYKPPSPS	60					
Db	1	MIRKSKITSVLSFCSRREIMWTILGRSALRELSQIAEALNKMRRLLEGSIYKPPSPS	60					
QY	61	SAEVRANKVAVSLKELGIRISKFGLEDBEQSVQLLOCTIOEYRGRTRDSVKTVLQDER	120					
Db	61	SAEVRANKVAVSLKELGIRISKFGLEDBEQSVQLLOCTIOEYRGRTRDSVKTVLQDER	120					
QY	121	QSQALLIKADIYYEERTCILRCVLLMTYFQDERHPRYRYEADCVMLKEELYSKTRQ	180					
Db	121	QSQALLIKADIYYEERTCILRCVLLMTYFQDERHPRYRYEADCVMLKEELYSKTRQ	180					
QY	181	FEELYTEAPRTWETHGNLMTEROYSRMFVQCLBESQMLBIIFLUYAFEMAPSDLLVLT	240					
Db	181	FEELYTEAPRTWETHGNLMTEROYSRMFVQCLBESQMLBIIFLUYAFEMAPSDLLVLT	240					
QY	241	KMFEEQSGFRQTNRHLVDETMQPVDRIGYFSALLIVEQMDIESLHKALDDBREHLQF	300					
Db	241	KMFEEQSGFRQTNRHLVDETMQPVDRIGYFSALLIVEQMDIESLHKALDDBREHLQF	300					
QY	301	AODGLICQMDCLMTFGDIPIHNAPVLLAWALLRHTLNPEETSVAVRKIGGTALQNLVQ	360					
Db	301	AODGLICQMDCLMTFGDIPIHNAPVLLAWALLRHTLNPEETSVAVRKIGGTALQNLVQ	360					
QY	361	YLRLRLQSLASGGNDCTSTACMCYGLSFPVLTSLTEHLTLGNQOIIDTRACEVLADPSL	420					
Db	361	YLRLRLQSLASGGNDCTSTACMCYGLSFPVLTSLTEHLTLGNQOIIDTRACEVLADPSL	420					
QY	421	PELFMGTEPTSGIGIILDSVCGMFPHLSPLOLRLALVSGKSTAKKVSFLLDMSPYNE	480					
Db	421	PELFMGTEPTSGIGIILDSVCGMFPHLSPLOLRLALVSGKSTAKKVSFLLDMSPYNE	480					
QY	481	LYHKRHVDYSHEDGTLMRQTPKLLYPLGGQTNLRIPOGTGQVMLDDBRAYLVWMEYSY	540					
Db	481	LYHKRHVDYSHEDGTLMRQTPKLLYPLGGQTNLRIPOGTGQVMLDDBRAYLVWMEYSY	540					
QY	541	SSWTLFTECEIEMLHVAVSTADVYOHQORXKPIIDLVHKVISTDLSADCLLPTSTRYML	600					
Db	541	SSWTLFTECEIEMLHVAVSTADVYOHQORXKPIIDLVHKVISTDLSADCLLPTSTRYML	600					
QY	601	LORLTYVISPVDVIVASVNCVLTVLAARNPAKVTDLRHTGFLPFAHVPVSSISQMSIAE	660					

```

Db 601 LQRLITTVISPPVDVIAACVNCITVLAARNPAKWTDLRHCFLEFVAHPVSSLSQMSAB 660
Qy 661 GMAAGYGNLIMNSEPOGEYGVTTAFLRLITTVKQGSSTQSGVPCVMFLKEMLP 720
Db 661 GMAAGYGNLIMNSEPOGEYGVTTAFLRLITTVKQGSSTQSGVPCVMFLKEMLP 720
Qy 721 SYHKWRVNSHVRQIGLITLIELIHAILINCHETDHSHPISLOPCTISLATTEAGOT 780
Db 721 SYHKWRVNSHVRQIGLITLIELIHAILINCHETDHSHPISLOPCTISLATTEAGOT 780
Qy 781 VINIMIGIVDTIDVWMAAOPRSDAEGQOGQOLIKTYKLAFTYNNVIRLKPSNVSP 840
Db 781 VINIMIGIVDTIDVWMAAOPRSDAEGQOGQOLIKTYKLAFTYNNVIRLKPSNVSP 840
Qy 841 LEQALSHGAGNNLIYAVAKYIYKHDPAIPRLAIQILKRLATVAPMSVYACIGNDAAA 900
Db 841 LEQALSHGAGNNLIYAVAKYIYKHDPAIPRLAIQILKRLATVAPMSVYACIGNDAAA 900
Qy 901 IRDAFLRLQSKIEDMRKMWILBFLTVAVETOPGLIELFLNLEVKQSDGSKESPJGMM 960
Db 901 IRDAFLRLQSKIEDMRKMWILBFLTVAVETOPGLIELFLNLEVKQSDGSKESPJGMM 960
Qy 961 SCLHAIVELIDSOQODRYWCPLHRAAIAELHALMODRDSAMLVLRTEKPFMENTLSP 1020
Db 961 SCLHAIVELIDSOQODRYWCPLHRAAIAELHALMODRDSAMLVLRTEKPFMENTLSP 1020
Qy 1021 LFGTLPSPSETSPSILETCALINKITICLETIYVVVKGSLDQSLKDTLKESIEKRPAYS 1080
Db 1021 LFGTLPSPSETSPSILETCALINKITICLETIYVVVKGSLDQSLKDTLKESIEKRPAYS 1080
Qy 1081 GYVSVLAVHVAETEGSSCTSLLEYOMLVSAWRMLIATTHADIMHLTDSVVRQLFDV 1140
Db 1081 GYVSVLAVHVAETEGSSCTSLLEYOMLVSAWRMLIATTHADIMHLTDSVVRQLFDV 1140
Qy 1141 LDGKALLLVASVNCILSGSMKCTLLILRQMKREISYDEILGPTEILEGVLQADQ 1200
Db 1141 LDGKALLLVASVNCILSGSMKCTLLILRQMKREISYDEILGPTEILEGVLQADQ 1200
Qy 1201 QLMKRTAKVSAFTTVLQMKEMKVSQDIPOYSQVLVANCETLQEBVIALPQTRHSLALG 1260
Db 1201 QLMKRTAKVSAFTTVLQMKEMKVSQDIPOYSQVLVANCETLQEBVIALPQTRHSLALG 1260
Qy 1261 SATEDKSMETDSCSRSHRDQRCVCLGLHAKELCEVEDDGSMVQVTRRLPIPTL 1320
Db 1261 SATEDKSMETDSCSRSHRDQRCVCLGLHAKELCEVEDDGSMVQVTRRLPIPTL 1320
Qy 1321 LTTLEVSLRMKONLHTEATLHLITTLARTQOGATAVAGAGITOSICLPLSVYQLSTNG 1380
Db 1321 LTTLEVSLRMKONLHTEATLHLITTLARTQOGATAVAGAGITOSICLPLSVYQLSTNG 1380
Qy 1381 TAQTPSASRSKSLDAPSPMGVYRLSMSLMEQLKTLRYNLFPEALDFVGHQERTLQCLNA 1440
Db 1381 TAQTPSASRSKSLDAPSPMGVYRLSMSLMEQLKTLRYNLFPEALDFVGHQERTLQCLNA 1440
Qy 1441 VRTVQSLACEBAHTVGFILQLSNPKEMKMHPLPOLMRDIOVNLGVYCOACTSLSHSRK 1500
Db 1441 VRTVQSLACEBAHTVGFILQLSNPKEMKMHPLPOLMRDIOVNLGVYCOACTSLSHSRK 1500
Qy 1501 MLQOYLONKNGDGLPSAQAORVORPPSAASAAPSSSKOPADTEASECOALHTVQYGLK 1560
Db 1501 MLQOYLONKNGDGLPSAQAORVORPPSAASAAPSSSKOPADTEASECOALHTVQYGLK 1560
Qy 1561 ILSTKTLAALRHFTPDVCOILIDQSLDLAEYNFLPALSTTPTPSEVAPSGTGLATVNV 1620
Db 1561 ILSTKTLAALRHFTPDVCOILIDQSLDLAEYNFLPALSTTPTPSEVAPSGTGLATVNV 1620
Qy 1621 ALNMLGELDKKKEPLTOAVGSTQAGERTLKSLMFEMENCFTYLLSQARVYLRDPVH 1680
Db 1621 ALNMLGELDKKKEPLTOAVGSTQAGERTLKSLMFEMENCFTYLLSQARVYLRDPVH 1680
Qy 1681 PRDKQRMKQELSELSTLSSLSRFRGAPSSPATGVLPQCKSTLSKASPESSOEPL 1740

```

```

Db 1681 PRDKQRMKQELSELSTLSSLSRFRGAPSSPATGVLPQCKSTLSKASPESSOEPL 1740
Qy 1741 IQLVQAFVRHMQR 1753
Db 1741 IQLVQAFVRHMQR 1753

RESULT 12
US-10-719-385-7
; Sequence 7, Application US/10719385
; Publication No. US20040209284A1
; GENERAL INFORMATION:
; APPLICANT: O'Toole et al.
; TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis
; FILE REFERENCE: 22058-582
; CURRENT FILING DATE: US/10/719,385
; PRIOR FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: PCT/US03/37339
; PRIOR FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: 60/4428,094
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 1753
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-385-7

Query Match 99.9%; Score 8999; DB 17; Length 1753;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1752; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MIRSKITTSVLSFCRSSRELTITLIGRSALRELSQIEALNKWRRLLEGISTYKPPSPS 60
Db 1 MIRSKITTSVLSFCRSSRELTITLIGRSALRELSQIEALNKWRRLLEGISTYKPPSPS 60
Qy 61 SAEKVRANKVDASPLKELGRLSKFGLGDEBSVOLQCYLOEYRGTGRDSVKTVLODER 120
Db 61 SAEKVRANKVDASPLKELGRLSKFGLGDEBSVOLQCYLOEYRGTGRDSVKTVLODER 120
Qy 121 QSQALILKIDVYEEERTCILRCVLAHLITTYFODERHPYRVEYACVDKLEKELVSKYRQ 180
Db 121 QSQALILKIDVYEEERTCILRCVLAHLITTYFODERHPYRVEYACVDKLEKELVSKYRQ 180
Qy 181 FEEELYKTEAPTWETHGNLMTEROVSQVOCLEQSMLEIIFLYAVFEMAPSDLVLT 240
Db 181 FEEELYKTEAPTWETHGNLMTEROVSQVOCLEQSMLEIIFLYAVFEMAPSDLVLT 240
Qy 241 KMFKEQGFSGQTRRHVDETMDFVDRIGFSAIILVEGMDISLHKCALDDRREIHOQ 300
Db 241 KMFKEQGFSGQTRRHVDETMDFVDRIGFSAIILVEGMDISLHKCALDDRREIHOQ 300
Qy 301 AODGLICQDMDCMLTFEGDIPHHAPVLLAWALLRHTLNPEBTSVVERKIGGTALQLVNQ 360
Db 301 AODGLICQDMDCMLTFEGDIPHHAPVLLAWALLRHTLNPEBTSVVERKIGGTALQLVNQ 360
Qy 361 YLTRILQSLASGGNDCTTSTACVCYGLLSFTVLSLEHTGNQODIITDCEVLADBSL 420
Db 361 YLTRILQSLASGGNDCTTSTACVCYGLLSFTVLSLEHTGNQODIITDCEVLADBSL 420
Qy 421 PELFMGTEPNSGIGIILDSVCGMPHLLSPLOLRLALVSGKSTAKKYYSFLDKMSFYNE 480
Db 421 PELFMGTEPNSGIGIILDSVCGMPHLLSPLOLRLALVSGKSTAKKYYSFLDKMSFYNE 480
Qy 481 LYKHKPDIIVSHEDGTLMRQTPRLIYPLGQTMRIPOGTVGQVMLDDRAVYLRWEYSY 540
Db 481 LYKHKPDIIVSHEDGTLMRQTPRLIYPLGQTMRIPOGTVGQVMLDDRAVYLRWEYSY 540
Qy 541 SSWTLFCEIEMLLHVSTADVIOHCORVKEIIDLVRKVISTDLSIADCLLPITSRIYML 600
Db 541 SSWTLFCEIEMLLHVSTADVIOHCORVKEIIDLVRKVISTDLSIADCLLPITSRIYML 600

```

QY 601 LQRLTTVISPVDVIAVCNCLTVLAARNPAKWTDLRHTGFLPFAVHVSLSQMIAS 660  
 DB 601 LQRLTTVISPVDVIAVCNCLTVLAARNPAKWTDLRHTGFLPFAVHVSLSQMIAS 660  
 QY 661 GNNAGGYGNLNLNSQPOGEVGTIAFLRLITLVKQGLQSTQSGVLPVCMFVKEMLP 720  
 DB 661 GNNAGGYGNLNLNSQPOGEVGTIAFLRLITLVKQGLQSTQSGVLPVCMFVKEMLP 720  
 QY 721 SYHKRNVSHGVREIGCILELHAAILNLCHEVTHSHSHTPSLOFLICSLATYEAQOT 780  
 DB 721 SYHKRNVSHGVREIGCILELHAAILNLCHEVTHSHSHTPSLOFLICSLATYEAQOT 780  
 QY 781 VININGIGVDTIDWMAAOPRSDGAEGQGGQLIKTVKLAFSVTNNVIRLKPSSNVSP 840  
 DB 781 VININGIGVDTIDWMAAOPRSDGAEGQGGQLIKTVKLAFSVTNNVIRLKPSSNVSP 840  
 QY 841 LEQALSQHGAGNNILAVLAKIYKHDPALPRALIQLLKRLATVAPWSVYACIGNDAAA 900  
 DB 841 LEQALSQHGAGNNILAVLAKIYKHDPALPRALIQLLKRLATVAPWSVYACIGNDAAA 900  
 QY 901 IRDAFLTLQSKIEDMRKIMLEFLVAVETOPGLIELFLNLEVKQSGDSKESFGMW 960  
 DB 901 IRDAFLTLQSKIEDMRKIMLEFLVAVETOPGLIELFLNLEVKQSGDSKESFGMW 960  
 QY 961 SCLHNAVIELDSQODRYWCPPLHRAAIAFLHMLMODRDSAMLVATKPKFWMENLTSP 1020  
 DB 961 SCLHNAVIELDSQODRYWCPPLHRAAIAFLHMLMODRDSAMLVATKPKFWMENLTSP 1020  
 QY 1021 LFGTLPSPSESEBSILETCALIMKILCLETIVYVKGSLDLSLDTLKKSIEKFAVWS 1080  
 DB 1021 LFGTLPSPSESEBSILETCALIMKILCLETIVYVKGSLDLSLDTLKKSIEKFAVWS 1080  
 QY 1081 GYVSLAVHVAETGSSCTSLLEYOMLVSAARMMLIATTHADIMHLDTSVVRQLFLDV 1140  
 DB 1081 GYVSLAVHVAETGSSCTSLLEYOMLVSAARMMLIATTHADIMHLDTSVVRQLFLDV 1140  
 QY 1141 LDGKRALLVPAVSNCLRLGSMKCTLLILRLKMKRELSGVDELGLPTEILBEGVADQ 1200  
 DB 1141 LDGKRALLVPAVSNCLRLGSMKCTLLILRLKMKRELSGVDELGLPTEILBEGVADQ 1200  
 QY 1201 QLMKRTAKVSAFITYLQMKEMKVSIPQYSQVLVNCETLOEVALPQPTRHSLALG 1260  
 DB 1201 QLMKRTAKVSAFITYLQMKEMKVSIPQYSQVLVNCETLOEVALPQPTRHSLALG 1260  
 QY 1261 SATEDKDSMETDSCSRSHRDQDQGVCVLGLHAKELCEVEDSDSWLOVTRRLPIPLTL 1320  
 DB 1261 SATEDKDSMETDSCSRSHRDQDQGVCVLGLHAKELCEVEDSDSWLOVTRRLPIPLTL 1320  
 QY 1321 LTTLEVSILRMKQNHFEATLHLLTLARTOOGATVAGAGITOSTICPLISVYQSTNG 1380  
 DB 1321 LTTLEVSILRMKQNHFEATLHLLTLARTOOGATVAGAGITOSTICPLISVYQSTNG 1380  
 QY 1381 TAQTPSASRSKSLDAPSPWGVYRLSMSLMBOULLKTLRLNPLPEALDFVGVHOBERTLOCLNA 1440  
 DB 1381 TAQTPSASRSKSLDAPSPWGVYRLSMSLMBOULLKTLRLNPLPEALDFVGVHOBERTLOCLNA 1440  
 QY 1441 VRTVQSLACLEADHTVGFILQLSNFMKEWHFHLPOLMRDIOVNLGYLCOACTSLHSRK 1500  
 DB 1441 VRTVQSLACLEADHTVGFILQLSNFMKEWHFHLPOLMRDIOVNLGYLCOACTSLHSRK 1500  
 QY 1501 MLQHYLONKNGDGLPSAVAOVQRPSPASASAPSSSQPADTEASQOALHTVVOYGLK 1560  
 DB 1501 MLQHYLONKNGDGLPSAVAOVQRPSPASASAPSSSQPADTEASQOALHTVVOYGLK 1560  
 QY 1561 ILSKTLAALRHFTPDVCOILLDOSLDLAENFLPALSFPTPTPSEVAPSFGLTALTANY 1620  
 DB 1561 ILSKTLAALRHFTPDVCOILLDOSLDLAENFLPALSFPTPTPSEVAPSFGLTALTANY 1620  
 QY 1621 ALNMLGELDKKKEPLTQAVGLSTQABSTRILKSLMTMENCYLLISQAMRYLRDPAYH 1680  
 DB 1621 ALNMLGELDKKKEPLTQAVGLSTQABSTRILKSLMTMENCYLLISQAMRYLRDPAYH 1680  
 QY 1681 PRDKQKQKQSLSELSLTLSSLSRYFRKGA PSSPATVLPSPQKSTLSLSKAPSESOEPL 1740

DB 1681 PRDKQKQKQSLSELSLTLSSLSRYFRKGA PSSPATVLPSPQKSTLSLSKAPSESOEPL 1740  
 QY 1741 IQLVQAFVRHMOR 1753  
 DB 1741 IQLVQAFVRHMOR 1753  
 RESULT 13  
 US-10-719-385-12  
 ; Sequence 12, Application US/10719385  
 ; Publication No. US20040209284A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: O'Toole et al.  
 ; TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis  
 ; FILE REFERENCE: 22058-582  
 ; CURRENT APPLICATION NUMBER: US/10/719,385  
 ; PRIOR FILING DATE: 2003-11-21  
 ; PRIOR APPLICATION NUMBER: PCT/US03/37339  
 ; PRIOR FILING DATE: 2003-11-21  
 ; PRIOR APPLICATION NUMBER: 60/428,094  
 ; NUMBER OF SEQ ID NOS: 26  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 12  
 ; LENGTH: 1753  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-10-719-385-12

Query Match 99.9%; Score 8999; DB 17; Length 1753;  
 Best local Similarity 99.9%; Pred. No. 0;  
 Matches 1752; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MIRSKITSVLSCRSRRLMTLLGRSALRELSQIEAEINRWRLLEGSLYKPPSPS 60  
 DB 1 MIRSKITSVLSCRSRRLMTLLGRSALRELSQIEAEINRWRLLEGSLYKPPSPS 60  
 QY 61 SAEKVANDVAPLDELGLRISKFLGDBEBSVOLQCYLOEDYVGTDSVTVQODR 120  
 DB 61 SAEKVANDVAPLDELGLRISKFLGDBEBSVOLQCYLOEDYVGTDSVTVQODR 120  
 QY 121 OSQALILKADYYEERTCILRCVHLTLTYFODERHPYVEVADCVKLEKELVSKYRQ 180  
 DB 121 OSQALILKADYYEERTCILRCVHLTLTYFODERHPYVEVADCVKLEKELVSKYRQ 180  
 QY 181 FEELYKTEAPYETHGNLMTERQVSNRFOCUREBSMLLEIFLYAYAFEMAPSDLLVLT 240  
 DB 181 FEELYKTEAPYETHGNLMTERQVSNRFOCUREBSMLLEIFLYAYAFEMAPSDLLVLT 240  
 QY 241 KMFKEGFSRQTNRLVDETMDPFVDRIGYSALILVGMDBESLHKCALDRRELHOF 300  
 DB 241 KMFKEGFSRQTNRLVDETMDPFVDRIGYSALILVGMDBESLHKCALDRRELHOF 300  
 QY 301 AODGLICQMDCLMTFGDI PHHAPVLLAMALLRHTLNEETS SVVRKIGTAIQLVNQ 360  
 DB 301 AODGLICQMDCLMTFGDI PHHAPVLLAMALLRHTLNEETS SVVRKIGTAIQLVNQ 360  
 QY 361 YLTRLQSLASGANDCTTSTACMCYVGLSFLVTSLEHTLNGNODIIDTACEVLADESL 420  
 DB 361 YLTRLQSLASGANDCTTSTACMCYVGLSFLVTSLEHTLNGNODIIDTACEVLADESL 420  
 QY 421 PELFWGTEPTSGILILDSVCGMFPHLSPLLQRLBALVSGSTAKKYSPFDKMSFYNE 480  
 DB 421 PELFWGTEPTSGILILDSVCGMFPHLSPLLQRLBALVSGSTAKKYSPFDKMSFYNE 480  
 QY 481 LYKHKPHDIVSHEDGTLMWRQTPKLLYPLGQGTNLRIPQGVQVMDRDAVLVMEVSY 540  
 DB 481 LYKHKPHDIVSHEDGTLMWRQTPKLLYPLGQGTNLRIPQGVQVMDRDAVLVMEVSY 540  
 QY 541 SSMWLTFCGIEMLLVVSTADVIQHCORVKKPIIDLHVHIVISTDLSIADCLLPITSRIYWL 600  
 DB 541 SSMWLTFCGIEMLLVVSTADVIQHCORVKKPIIDLHVHIVISTDLSIADCLLPITSRIYWL 600

```

QY 601 LQRLITTVISPVVDYIASCVCNCLTVLAARNPAKWTDLHRTGFLPVAVHSPVSLSOMISAE 660
Db 601 LQRLITTVISPVVDYIASCVCNCLTVLAARNPAKWTDLHRTGFLPVAVHSPVSLSOMISAE 660
QY 661 GMAAGYGNLMLMNEBQPOGEYGVITIAFLRLITTVKQGLSGTQSGVLPCVMFVLEKMLP 720
Db 661 GMAAGYGNLMLMNEBQPOGEYGVITIAFLRLITTVKQGLSGTQSGVLPCVMFVLEKMLP 720
QY 721 SYHKWRNHSVGRQIGCLIELIHALINLCHETDHSHTPSLOFLCISLATTEAGOT 780
Db 721 SYHKWRNHSVGRQIGCLIELIHALINLCHETDHSHTPSLOFLCISLATTEAGOT 780
QY 781 VINIMGIGVDTIIDWMAAPRSDGEGGQGLIKTYKLAFSVTNNVIRLKPSPNVSP 840
Db 781 VINIMGIGVDTIIDWMAAPRSDGEGGQGLIKTYKLAFSVTNNVIRLKPSPNVSP 840
QY 841 LEQALSOHGAHGNMLIAVLAKYIYHKGDPALPRLAIQLKSLATVAPMSVYACIGNDAAA 900
Db 841 LEQALSOHGAHGNMLIAVLAKYIYHKGDPALPRLAIQLKSLATVAPMSVYACIGNDAAA 900
QY 901 IRDAFLTRLQSKIEDMRKVMILEFLYAVETOPGLIELFLNLEVKQSDGSKESFLTMW 960
Db 901 IRDAFLTRLQSKIEDMRKVMILEFLYAVETOPGLIELFLNLEVKQSDGSKESFLTMW 960
QY 961 SCLHAEVLIDSOODRWCPEPLHRAIAFLHLMODRDSAMLVLRTPKPMENLTSP 1020
Db 961 SCLHAEVLIDSOODRWCPEPLHRAIAFLHLMODRDSAMLVLRTPKPMENLTSP 1020
QY 1021 LFGTLSPSESTSEPSILETCALIMKICLEIYVVVKGSLDQSLKDTLKFSIEKRFAYMS 1080
Db 1021 LFGTLSPSESTSEPSILETCALIMKICLEIYVVVKGSLDQSLKDTLKFSIEKRFAYMS 1080
QY 1081 GYVVSIAVHVAETBESSCTSLLEYOMLVSAARMILITTHADIMHLDTSVVRQLFDV 1140
Db 1081 GYVVSIAVHVAETBESSCTSLLEYOMLVSAARMILITTHADIMHLDTSVVRQLFDV 1140
QY 1141 LDGFKALLVPASVNCRLGSMKCTLLILRLKQRELSYDEILGPTLEILEGVLQDQ 1200
Db 1141 LDGFKALLVPASVNCRLGSMKCTLLILRLKQRELSYDEILGPTLEILEGVLQDQ 1200
QY 1201 QLMEXTKAVFSAFITVLQMKEMKVSIPQYSQLVLANCETLQEBEVALPQTRSHLALG 1260
Db 1201 QLMEXTKAVFSAFITVLQMKEMKVSIPQYSQLVLANCETLQEBEVALPQTRSHLALG 1260
QY 1261 SATEDKOSMETDDCSRRHRODQGVCLGLHAKELCEVEDGDGSMQVTRRLPIPLTL 1320
Db 1261 SATEDKOSMETDDCSRRHRODQGVCLGLHAKELCEVEDGDGSMQVTRRLPIPLTL 1320
QY 1321 LTTLEVSILRMKONLHFTBATLHLTLTLARTOOGAFAVAGAGITOSICPLISVYQLSTNG 1380
Db 1321 LTTLEVSILRMKONLHFTBATLHLTLTLARTOOGAFAVAGAGITOSICPLISVYQLSTNG 1380
QY 1381 TAQTPSASRKSILDAFSPGVRRLSMLBOLKTLRYNPLPEALDFVGHQERTLOCLANA 1440
Db 1381 TAQTPSASRKSILDAFSPGVRRLSMLBOLKTLRYNPLPEALDFVGHQERTLOCLANA 1440
QY 1441 VRTVQSLACEBADHTVGPILQLSNFMKEWHPHLFQMKRDIOVNIQYICQACTSLHSRK 1500
Db 1441 VRTVQSLACEBADHTVGPILQLSNFMKEWHPHLFQMKRDIOVNIQYICQACTSLHSRK 1500
QY 1501 MLQHYLQKNGDGPBSAFAORVORPSPASASAPSSKOPADTESEBOALHTVQYGLK 1560
Db 1501 MLQHYLQKNGDGPBSAFAORVORPSPASASAPSSKOPADTESEBOALHTVQYGLK 1560
QY 1561 ILSTKLALRHFTPDVCOILLDQSDJDLAEYNFLFALSTFTTTPDSEVAPSGETLLATVNV 1620
Db 1561 ILSTKLALRHFTPDVCOILLDQSDJDLAEYNFLFALSTFTTTPDSEVAPSGETLLATVNV 1620
QY 1621 ALNMLGELDKKKEPLTQAVGLSTQAEGRITLKSILMTMENC FYLLISQAMRYLRDPVH 1680
Db 1621 ALNMLGELDKKKEPLTQAVGLSTQAEGRITLKSILMTMENC FYLLISQAMRYLRDPVH 1680

```

```

QY 1681 PRDKORMKOBLSSELSTLSSLSRYFRRGAPSSPATGVLPSPOGKSTLSKASPEBOEPL 1740
Db 1681 PRDKORMKOBLSSELSTLSSLSRYFRRGAPSSPATGVLPSPOGKSTLSKASPEBOEPL 1740
QY 1741 IOLVOAFVRHMQR 1753
Db 1741 IOLVOAFVRHMQR 1753

RESULT 14
US-10-719-385-18
; Sequence 18, Application US/10719385
; Publication No. US20040209284A1
; GENERAL INFORMATION:
; APPLICANT: O'Toole et al.
; TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis
; FILE REFERENCE: 22058-582
; CURRENT APPLICATION NUMBER: US/10/719,385
; PRIOR FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: PCT/US03/37339
; PRIOR FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: 60/428,094
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 18
; LENGTH: 1753
; TYPE: PRN
; ORGANISM: Homo sapiens
US-10-719-385-18

Query Match 99.9%; Score 8998; DB 17; Length 1753;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1751; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MIRSKITSVLSFCRSSHEMTIILGSAARELSQIEAEINKMWRLLLEGSLYKPPSPS 60
Db 1 MIRSKITSVLSFCRSSHEMTIILGSAARELSQIEAEINKMWRLLLEGSLYKPPSPS 60
QY 61 SAEYKANKQVAPSLKELGLRISKFLGIDBQSVQLQCTYQOEYRGTRDSVKTVLODER 120
Db 61 SAEYKANKQVAPSLKELGLRISKFLGIDBQSVQLQCTYQOEYRGTRDSVKTVLODER 120
QY 121 QSOALLIKIADYYEERTCIRCVLHLLTYQODERHPRVEYACVDLKELELSKYRQ 180
Db 121 QSOALLIKIADYYEERTCIRCVLHLLTYQODERHPRVEYACVDLKELELSKYRQ 180
QY 181 FEELYKTEAPTWETHGNLMTEROVSRMFVQCLRBQSMLEIIPYYAYFENAPSDLVLT 240
Db 181 FEELYKTEAPTWETHGNLMTEROVSRMFVQCLRBQSMLEIIPYYAYFENAPSDLVLT 240
QY 241 KMFEKQGFSGQTRNRLVDETMDFVDRIGFSALLIVEGMDISLHKCALDDRRRLHQF 300
Db 241 KMFEKQGFSGQTRNRLVDETMDFVDRIGFSALLIVEGMDISLHKCALDDRRRLHQF 300
QY 301 AODGLICODMDCMLTREGDIPRHA PVLAMALBHTLNPEETSSVVRKIGGTALQNAVQ 360
Db 301 AODGLICODMDCMLTREGDIPRHA PVLAMALBHTLNPEETSSVVRKIGGTALQNAVQ 360
QY 361 YLTRLLQSLASGNDCTTSTACMKCVYGLSFTLSLEHTLGNQODIIDTACEVLADPSL 420
Db 361 YLTRLLQSLASGNDCTTSTACMKCVYGLSFTLSLEHTLGNQODIIDTACEVLADPSL 420
QY 421 PELFWGTEPTSGLGIIIDSVCGMFPPLISPLQLLRALVSGKSTAKKYVSLDKKSFNE 480
Db 421 PELFWGTEPTSGLGIIIDSVCGMFPPLISPLQLLRALVSGKSTAKKYVSLDKKSFNE 480
QY 481 LYKXKPHDVISHEDGTLMBRQTPKLVPLGGQTNLRIPQGTVGQVMLDDBRAYLVKMEYSY 540
Db 481 LYKXKPHDVISHEDGTLMBRQTPKLVPLGGQTNLRIPQGTVGQVMLDDBRAYLVKMEYSY 540
QY 541 SSWTLFCEIBMLLHVSTADVIOHCORVKPIIDLWVKVISTDLSIADCLLPITSRIYML 600

```

Db 541 SSWTLFTCEIEMLHVSTADVIQHQRVKPIIDLHVHVIISTDLSIADCLPITSRIYML 600  
 Qy 601 LQRLTTVISPVDVYASCVNCLTVLAARNPAKWTDLRHTEGFLPVAHPVSSLQMSISAE 660  
 Db 601 LQRLTTVISPVDVYASCVNCLTVLAARNPAKWTDLRHTEGFLPVAHPVSSLQMSISAE 660  
 Qy 661 GNNAGYGNLNMNSPOEOEVYTAFLRLITTLVKGOLGOSQSGCLPCWCFVTKXELP 720  
 Db 661 GNNAGYGNLNMNSPOEOEVYTAFLRLITTLVKGOLGOSQSGCLPCWCFVTKXELP 720  
 Qy 721 SYHKRNVSHGVEQIGCLILELHAIIINLCHETDLHSHSTPSLOFLCISLAYTEAGOT 780  
 Db 721 SYHKRNVSHGVEQIGCLILELHAIIINLCHETDLHSHSTPSLOFLCISLAYTEAGOT 780  
 Qy 781 VINIMGIVDTIDMMAOPSDGAEQGGGOLLKTKYKLAFASTYNNITRLKPSNVSP 840  
 Db 781 VINIMGIVDTIDMMAOPSDGAEQGGGOLLKTKYKLAFASTYNNITRLKPSNVSP 840  
 Qy 841 LEQALSHQAGHNNLIAYLAKYIYKHDPALPRLAIQLKRLATVAPMSVYACIGNDAAA 900  
 Db 841 LEQALSHQAGHNNLIAYLAKYIYKHDPALPRLAIQLKRLATVAPMSVYACIGNDAAA 900  
 Qy 901 IRDAFLTLQSKIEIMRIKWMLEPLVAVETOPGLILFLNLEVKGSDGSKPSIGMW 960  
 Db 901 IRDAFLTLQSKIEIMRIKWMLEPLVAVETOPGLILFLNLEVKGSDGSKPSIGMW 960  
 Qy 961 SCLHVLIELISQOQDRYCPPLHRAAIAFLHLMODRDSAMLVJTRKPMENITSP 1020  
 Db 961 SCLHVLIELISQOQDRYCPPLHRAAIAFLHLMODRDSAMLVJTRKPMENITSP 1020  
 Qy 1021 LFGTISPSETSEPSILETCALIMKIICLEIYVYVKSJLDSLOTLKKSIEKRFAYMS 1080  
 Db 1021 LFGTISPSETSEPSILETCALIMKIICLEIYVYVKSJLDSLOTLKKSIEKRFAYMS 1080  
 Qy 1081 GYVSLAHVAETGSSCTSLLEYQMLVSAMRMIIITTTADIMHLDTSVVRQLELDV 1140  
 Db 1081 GYVSLAHVAETGSSCTSLLEYQMLVSAMRMIIITTTADIMHLDTSVVRQLELDV 1140  
 Qy 1141 LDGTRKALLVPASVNCRLGSMKCTLLILRLKQKRELSVDEILGPTLIEGVLAQDQ 1200  
 Db 1141 LDGTRKALLVPASVNCRLGSMKCTLLILRLKQKRELSVDEILGPTLIEGVLAQDQ 1200  
 Qy 1201 QLMKRTAKYVSAITTVLQMKEMKVSIPQYSQVLAVNCEVTLQOEVIAPLQDTRHSIALG 1260  
 Db 1201 QLMKRTAKYVSAITTVLQMKEMKVSIPQYSQVLAVNCEVTLQOEVIAPLQDTRHSIALG 1260  
 Qy 1261 SATEDXDSMETDSCSRSHRDQDQVCYLGLHLAKECEVEDSDSWLQVTRRLPILPTL 1320  
 Db 1261 SATEDXDSMETDSCSRSHRDQDQVCYLGLHLAKECEVEDSDSWLQVTRRLPILPTL 1320  
 Qy 1321 LTTLEVSILRMKONLHFEATLHLTLARTQOGATAVAGAGITQOSICLPILSYVQJSTNG 1380  
 Db 1321 LTTLEVSILRMKONLHFEATLHLTLARTQOGATAVAGAGITQOSICLPILSYVQJSTNG 1380  
 Qy 1381 TAOPPSASRSKSLDAPSPRGVYRLSMSLMEQLKTLARNFLPEALDFGVHQBERTLOCLNA 1440  
 Db 1381 TAOPPSASRSKSLDAPSPRGVYRLSMSLMEQLKTLARNFLPEALDFGVHQBERTLOCLNA 1440  
 Qy 1441 VRTVQSLACLEADHTVGFILQSLNFMKEWHFHLPOLMRDIOVNLGYLCOACTSLHSRK 1500  
 Db 1441 VRTVQSLACLEADHTVGFILQSLNFMKEWHFHLPOLMRDIOVNLGYLCOACTSLHSRK 1500  
 Qy 1501 MLOHYLONKNGDGLPSAVAOQVORPPSAASAAPSSSQPAADTEASBOALHTVOYGLK 1560  
 Db 1501 MLOHYLONKNGDGLPSAVAOQVORPPSAASAAPSSSQPAADTEASBOALHTVOYGLK 1560  
 Qy 1561 ILSTKTLAALHFPDVOOIIILDOSLDLAEYNELPALSFPTPTDSEVABSFGLTATVNY 1620  
 Db 1561 ILSTKTLAALHFPDVOOIIILDOSLDLAEYNELPALSFPTPTDSEVABSFGLTATVNY 1620  
 Qy 1621 ALNMLGELDKKKEBPLTQAVGLSTQABSTRILKSLMTFMENCFFYLISQAMRYLRDPAYH 1680  
 Db 1621 ALNMLGELDKKKEBPLTQAVGLSTQABSTRILKSLMTFMENCFFYLISQAMRYLRDPAYH 1680

Qy 1681 PRDKQKQKOBLSSELSTLSSLSRFFRGAPSSPATGVLPSPOCKSTSLSKASPEQOEP 1740  
 Db 1681 PRDKQKQKOBLSSELSTLSSLSRFFRGAPSSPATGVLPSPOCKSTSLSKASPEQOEP 1740  
 Qy 1741 IOLVQAFVRHMOR 1753  
 Db 1741 IOLVQAFVRHMOR 1753  
 RESULT 15  
 US-10-719-385-9  
 ; Sequence 9, Application US/10719385  
 ; Publication No. US20040209284A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: O'Toole et al.  
 ; TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis  
 ; FILE REFERENCE: 22058-582  
 ; CURRENT APPLICATION NUMBER: US/10/719,385  
 ; PRIOR FILING DATE: 2003-11-21  
 ; PRIOR APPLICATION NUMBER: PCT/US03/37339  
 ; PRIOR FILING DATE: 2003-11-21  
 ; PRIOR APPLICATION NUMBER: 60/428,094  
 ; NUMBER OF SEQ ID NOS: 26  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 9  
 ; LENGTH: 1753  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-10-719-385-9  
 Query Match 99.9%; Score 8997; DB 17; Length 1753;  
 Best Local Similarity 99.9%; Pred. No. 0;  
 Matches 1752; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 Qy 1 MIRSKITSVLSFCRASSREIMTTLGRSALRELSQIEAEINKMRRLBGLSYKPPSP 60  
 Db 1 MIRSKITSVLSFCRASSREIMTTLGRSALRELSQIEAEINKMRRLBGLSYKPPSP 60  
 Qy 61 SAEKYANKDVASPLKELGIRISKFLGDBEVSQVLLQCYLOEDYRGTRDSVXTVLQDER 120  
 Db 61 SAEKYANKDVASPLKELGIRISKFLGDBEVSQVLLQCYLOEDYRGTRDSVXTVLQDER 120  
 Qy 121 QSOALILKTDADYBERTCILRCVHLTLVPODERHPVVEYADCVDKLEKELVSKYRQ 180  
 Db 121 QSOALILKTDADYBERTCILRCVHLTLVPODERHPVVEYADCVDKLEKELVSKYRQ 180  
 Qy 181 FEELYTEAPVETHTGNLMTERQVSRWFVQCLREOSMLLEIIFLYAYEFEMAPSDLLVLT 240  
 Db 181 FEELYTEAPVETHTGNLMTERQVSRWFVQCLREOSMLLEIIFLYAYEFEMAPSDLLVLT 240  
 Qy 241 KMFKEGFSRQTNRLVDETHDPFVDRIGYSALILVEGMDIESLHKCALDDREILHOF 300  
 Db 241 KMFKEGFSRQTNRLVDETHDPFVDRIGYSALILVEGMDIESLHKCALDDREILHOF 300  
 Qy 301 AODGLICQDMDCMLTFGDI PHHAPVYLAWALLRHTLANEETSSVYRKIGGTAICQNVPO 360  
 Db 301 AODGLICQDMDCMLTFGDI PHHAPVYLAWALLRHTLANEETSSVYRKIGGTAICQNVPO 360  
 Qy 361 YLTRLIQSLASGANDCTTSTACMCVYGLISFVLSLEHTLGNODIIDTACEVLADPSL 420  
 Db 361 YLTRLIQSLASGANDCTTSTACMCVYGLISFVLSLEHTLGNODIIDTACEVLADPSL 420  
 Qy 421 PELFWGTEPTSGIITLDSVCGMFPHLSPILQLRALVSGSTAKYVSFLDKRSFYNE 480  
 Db 421 PELFWGTEPTSGIITLDSVCGMFPHLSPILQLRALVSGSTAKYVSFLDKRSFYNE 480  
 Qy 481 LYKHKPHDIVISHEGDTLMRQTPKLLYPLGQTNLRIPQYGVGYMLDDRAVLVMEYSY 540  
 Db 481 LYKHKPHDIVISHEGDTLMRQTPKLLYPLGQTNLRIPQYGVGYMLDDRAVLVMEYSY 540  
 Qy 541 SSWTLFTCEIEMLHVSTADVIQHQRVKPIIDLHVHVIISTDLSIADCLPITSRIYML 600

```
Db 541 SSWLFLCEIEMLAHVSTADVIOHCQVRKPIIDLVKHAVISTDLSIDCLLPISRIYML 600
Qy 601 LQRLITVISPVDVIVASCVNCLTVLAARNPAKWTDLRHGFLFPVAHPVSSLSQMSAB 660
Db 601 LQRLITVISPVDVIVASCVNCLTVLAARNPAKVTDLRHGFLFPVAHPVSSLSQMSAB 660
Qy 661 GMAAGGVNLLMSEOSOGEGVGTIAFLRLITLVKQOLGTSOGIVPCMFVLMKMLP 720
Db 661 GMAAGGVNLLMSEOSOGEGVGTIAFLRLITLVKQOLGTSOGIVPCMFVLMKMLP 720
Qy 721 SYHKMYNSHGVREIQGLILELIIHALINLCHETDLSSHSPISQFLICSLAYTEAGOT 780
Db 721 SYHKMYNSHGVREIQGLILELIIHALINLCHETDLSSHSPISQFLICSLAYTEAGOT 780
Qy 781 VINIMIGVDTIDVMAAOPRSDAEGOGQOLIKVYKLAFSVTNNVIRLKPSSNVSP 840
Db 781 VINIMIGVDTIDVMAAOPRSDAEGOGQOLIKVYKLAFSVTNNVIRLKPSSNVSP 840
Qy 841 LEOALSGHAGNNLIVAKYIYHKDPAIPLAIDOLKRLATVAPMSVYACLGNDAAA 900
Db 841 LEOALSGHAGNNLIVAKYIYHKDPAIPLAIDOLKRLATVAPMSVYACLGNDAAA 900
Qy 901 IRDAFLTRLOSIEDMRKIMWILEFLTVAVETOPGLIELFLNLEVKSGDSKESFSGMW 960
Db 901 IRDAFLTRLOSIEDMRKIMWILEFLTVAVETOPGLIELFLNLEVKSGDSKESFSGMW 960
Qy 961 SCHAVLELIDSOQODRYWCPLLHRAAIAFLAALMODRDSAMLVLRKPKFMENTLSP 1020
Db 961 SCHAVLELIDSOQODRYWCPLLHRAAIAFLAALMODRDSAMLVLRKPKFMENTLSP 1020
Qy 1021 LFGTLSPSESESESIETCLIMKIICLETYYVVKSLDOSTLMDTLKKSEIERFVWS 1080
Db 1021 LFGTLSPSESESESIETCLIMKIICLETYYVVKSLDOSTLMDTLKKSEIERFVWS 1080
Qy 1081 GYVXSLAVHVAETEGSSCTSLLEYQMLVSAMRMILLIATTTADIMHLDTSVVRQLFLDV 1140
Db 1081 GYVXSLAVHVAETEGSSCTSLLEYQMLVSAMRMILLIATTTADIMHLDTSVVRQLFLDV 1140
Qy 1141 LDGTRALLLVASVNCRLSGSMKCTLLIILIRKMRKEISVDEILGPTLEILEVQLADQ 1200
Db 1141 LDGTRALLLVASVNCRLSGSMKCTLLIILIRKMRKEISVDEILGPTLEILEVQLADQ 1200
Qy 1201 QLMKTKAKVPSAIIYVLQMKEMKVSIDIPOYSQVLANVCEFLQREVALPQOTRHSIALG 1260
Db 1201 QLMKTKAKVPSAIIYVLQMKEMKVSIDIPOYSQVLANVCEFLQREVALPQOTRHSIALG 1260
Qy 1261 SATEDKSMETDSCSRSHRQDQGVCVLGLHLAKELCEVEDDGSMLQVTRRLPIPLTL 1320
Db 1261 SATEDKSMETDSCSRSHRQDQGVCVLGLHLAKELCEVEDDGSMLQVTRRLPIPLTL 1320
Qy 1321 LTTLEVSIRMKONLHFTETATLHLLTLARTOGATAVAGAGITOSICLPLISVYQLSTNG 1380
Db 1321 LTTLEVSIRMKONLHFTETATLHLLTLARTOGATAVAGAGITOSICLPLISVYQLSTNG 1380
Qy 1381 TAQCPSSSRSLDAPSPWGVVRLSMSLMEQLKTLRVNLFPEALDPGVGHERLQCLANA 1440
Db 1381 TAQCPSSSRSLDAPSPWGVVRLSMSLMEQLKTLRVNLFPEALDPGVGHERLQCLANA 1440
Qy 1441 VRTVQSLACEAETHVGFILQLSNFMKEWHFHLPOLMRDIQVNLGYLCOACTSLHSRK 1500
Db 1441 VRTVQSLACEAETHVGFILQLSNFMKEWHFHLPOLMRDIQVNLGYLCOACTSLHSRK 1500
Qy 1501 MLQHYLONKNGDGLPSAVAQRFVQRPSSAASAPSSSKQPADTEASEQOALHTVQYGLK 1560
Db 1501 MLQHYLONKNGDGLPSAVAQRFVQRPSSAASAPSSSKQPADTEASEQOALHTVQYGLK 1560
Qy 1561 ILSTTLAALHRTFDVQIILLDOSIDLAEVNFALSTTTTPPSEVAPASGCTILLATNV 1620
Db 1561 ILSTTLAALHRTFDVQIILLDOSIDLAEVNFALSTTTTPPSEVAPASGCTILLATNV 1620
Qy 1621 ALNNMGLDKKEPLTOAVGLSTQAEGRTLKSLMFTMENC FYLLISQAMRYLDRPAVH 1680
```

```
Db 1621 ALNNMGLDKKEPLTOAVGLSTQAEGRTLKSLMFTMENC FYLLISQAMRYLDRPAVH 1680
Qy 1681 PRDKQRMKQELSSSLSTLSSLSRYFRRGAPSSPATGVLPBPOCKSTLSKASPESEQEPL 1740
Db 1681 PRDKQRMKQELSSSLSTLSSLSRYFRRGAPSSPATGVLPBPOCKSTLSKASPESEQEPL 1740
Qy 1741 IOLVQAFVRHMOR 1753
Db 1741 IOLVQAFVRHMOR 1753

RESULT 16
US-10-719-385-17
; Sequence 17, Application US/10719385
; Publication No. US20040209284A1
; GENERAL INFORMATION:
; APPLICANT: O'toole et al.
; TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis
; FILE REFERENCE: 22058-582
; CURRENT APPLICATION NUMBER: US/10/719,385
; PRIOR FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: PCT/US03/37339
; PRIOR FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: 60/428,094
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 17
; LENGTH: 1753
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-385-17

Query Match 99.8%; Score 8993; DB 17; Length 1753;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1751; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MIRSKITSVLSPCRSSREIMTILGRSALREISOIEAEINKHMRRLLEGISYKPPSPS 60
Db 1 MIRSKITSVLSPCRSSREIMTILGRSALREISOIEAEINKHMRRLLEGISYKPPSPS 60
Qy 61 SAEKYKANKVAPSLKELGLIRISKFLGLDEBOSVOLQCYLOEDYRGTRSVKTVLODER 120
Db 61 SAEKYKANKVAPSLKELGLIRISKFLGLDEBOSVOLQCYLOEDYRGTRSVKTVLODER 120
Qy 121 QSQALILKADYYEERTCILRCVHLHTYFQDERHRYRVEYADCVKLEKELVSKTRQ 180
Db 121 QSQALILKADYYEERTCILRCVHLHTYFQDERHRYRVEYADCVKLEKELVSKTRQ 180
Qy 181 FEEIYKTEAPWETHGNLMTEROYSRWFOCLRQSMLEIIEIYVAYFENAPBDLVLT 240
Db 181 FEEIYKTEAPWETHGNLMTEROYSRWFOCLRQSMLEIIEIYVAYFENAPBDLVLT 240
Qy 241 KMFKEQFGSROTNRHLVDETMDFPVNRIGYFSALIIVEGMDISLHKCALDDRREILHQF 300
Db 241 KMFKEQFGSROTNRHLVDETMDFPVNRIGYFSALIIVEGMDISLHKCALDDRREILHQF 300
Qy 301 AODGLICQDMDCMLTFGDIPHAPVLLAWALLRHTLNPEETSIVVRKIGGTALQLVNFQ 360
Db 301 AODGLICQDMDCMLTFGDIPHAPVLLAWALLRHTLNPEETSIVVRKIGGTALQLVNFQ 360
Qy 361 YLTRLLOSLSGSGNDCTTSTACMKCVGLSLFVLSLEHLTLGNODIITDACEVLADPSL 420
Db 361 YLTRLLOSLSGSGNDCTTSTACMKCVGLSLFVLSLEHLTLGNODIITDACEVLADPSL 420
Qy 421 PELFWGTEPTSGLGIILDSVCGMFPHLSPILLQRLALVSGSKSTAKKYVSFLDRMSFYNE 480
Db 421 PELFWGTEPTSGLGIILDSVCGMFPHLSPILLQRLALVSGSKSTAKKYVSFLDRMSFYNE 480
Qy 481 LYKXKPHDIVISHEDGTLMRQTPKLLYPLGQTNLRIPQGTVGQVMLDRAYLVREMSY 540
Db 481 LYKXKPHDIVISHEDGTLMRQTPKLLYPLGQTNLRIPQGTVGQVMLDRAYLVREMSY 540
```



```

QY 541 SSWTLFCEIEMTLHVSTADYIOHCRVKPIIDLHVKVISINDLSIADCLLPITSRIYML 600
DB 541 SSWTLFCEIEMTLHVSTADYIOHCRVKPIIDLHVKVISINDLSIADCLLPITSRIYML 600
QY 601 LQRLTTVISPVDVIASCNCCLTVLAARNPAKWTDLHRTGFLPFAHVPVSLSQMSAE 660
DB 601 LQRLTTVISPVDVIASCNCCLTVLAARNPAKWTDLHRTGFLPFAHVPVSLSQMSAE 660
QY 661 GMMNGGVNLLMNSBQGEYGVTAFLRLITTLVKGQSGTSGQLVPCWPFYKEMLP 720
DB 661 GMMNGGVNLLMNSBQGEYGVTAFLRLITTLVKGQSGTSGQLVPCWPFYKEMLP 720
QY 721 SYHKRVNHSVREGIQLIELIHLAHLNCHETDHSHTPSLOFLCTISLAYTEAGOT 780
DB 721 SYHKRVNHSVREGIQLIELIHLAHLNCHETDHSHTPSLOFLCTISLAYTEAGOT 780
QY 781 VININGICVDTIDWMAAOPRSDGAEQGGQGLIKTVKLAFSVTNNVIRLKPSNVSP 840
DB 781 VININGICVDTIDWMAAOPRSDGAEQGGQGLIKTVKLAFSVTNNVIRLKPSNVSP 840
QY 841 LEQALSOHGAHNNLIATLAKYIYHKHDPALPRALIQILKRLATVAPMSVYACIGNDAAA 900
DB 841 LEQALSOHGAHNNLIATLAKYIYHKHDPALPRALIQILKRLATVAPMSVYACIGNDAAA 900
QY 901 IRDAFLTRLOSKIEDMRKIMLEFLVAVETOPGLEFLNLEVKQSDGSKFESIGMW 960
DB 901 IRDAFLTRLOSKIEDMRKIMLEFLVAVETOPGLEFLNLEVKQSDGSKFESIGMW 960
QY 961 SCILAVBELIDSQOQDRWCPLHRAAIAFLAHLMODRDSAMVLRTRKPEMENTTSP 1020
DB 961 SCILAVBELIDSQOQDRWCPLHRAAIAFLAHLMODRDSAMVLRTRKPEMENTTSP 1020
QY 1021 LFGTLSPSETSBSILETALIMKIICLEIYVYVKSLSQSLDQTLKKSIEKRPYWS 1080
DB 1021 LFGTLSPSETSBSILETALIMKIICLEIYVYVKSLSQSLDQTLKKSIEKRPYWS 1080
QY 1081 GYVKSIAVHVAETEGSSCTSLLEYQMLVSAVRMLLIATTHADIMHLDTSVVRQLFLDV 1140
DB 1081 GYVKSIAVHVAETEGSSCTSLLEYQMLVSAVRMLLIATTHADIMHLDTSVVRQLFLDV 1140
QY 1141 LDGKALLVPAVNCRLGSMKCTLLILIRKOKREISYDELIGPTEILEBVLOAQD 1200
DB 1141 LDGKALLVPAVNCRLGSMKCTLLILIRKOKREISYDELIGPTEILEBVLOAQD 1200
QY 1201 QLMKRTAKYFSAITVLOMKEMKVSIDIPQYSQVLVNCETLOEVIALLFQTSHTALG 1260
DB 1201 QLMKRTAKYFSAITVLOMKEMKVSIDIPQYSQVLVNCETLOEVIALLFQTSHTALG 1260
QY 1261 SATEDKQSMETDSCSRHRDQDQDVCVGLGHLAKELCEVEDGDSWLQVTRRLPILPTL 1320
DB 1261 SATEDKQSMETDSCSRHRDQDQDVCVGLGHLAKELCEVEDGDSWLQVTRRLPILPTL 1320
QY 1321 LTTLEVLKMKONHFEATLHLLTLARTOQGAATAAGITQSLCPILSYVQLSTNG 1380
DB 1321 LTTLEVLKMKONHFEATLHLLTLARTOQGAATAAGITQSLCPILSYVQLSTNG 1380
QY 1381 TAQTPSASRSKSLDAPSWPGVYRLSMSIMEQLIKTLRYNFLPEALDFGVHQBERTLOCLNA 1440
DB 1381 TAQTPSASRSKSLDAPSWPGVYRLSMSIMEQLIKTLRYNFLPEALDFGVHQBERTLOCLNA 1440
QY 1441 VRTVQSLACLEADHTVGFILQSLNFKEMWHFHLPOLMRDIOQVNLGYLCAQCTSLHSRK 1500
DB 1441 VRTVQSLACLEADHTVGFILQSLNFKEMWHFHLPOLMRDIOQVNLGYLCAQCTSLHSRK 1500
QY 1501 MLOHYLONKKGDDGSAVAORVORPSPAASAAPSSKOPADTEASQOALHTVOYGLK 1560
DB 1501 MLOHYLONKKGDDGSAVAORVORPSPAASAAPSSKOPADTEASQOALHTVOYGLK 1560
QY 1561 ILSKTLAALHNPDPVCOILIDSLDAEYNFLPALSFTTPTFDESEVAPSPFGTLLATVNY 1620
DB 1561 ILSKTLAALHNPDPVCOILIDSLDAEYNFLPALSFTTPTFDESEVAPSPFGTLLATVNY 1620
QY 1621 ALNMLGELDKKKEPLTQAVGLSTQABGTRTLKSLMFTMENCFYLLISQAMRYLRDPVH 1680

```

```

DB 1621 ALNMLGELDKKKEPLTQAVGLSTQABGTRTLKSLMFTMENCFYLLISQAMRYLRDPVH 1680
QY 1681 PRDKQRMKOBELISELSTLSLSRYFRRGAPSSPATGVLPSPQKSTLSKASPEQOPL 1740
DB 1681 PRDKQRMKOBELISELSTLSLSRYFRRGAPSSPATGVLPSPQKSTLSKASPEQOPL 1740
QY 1741 IQLVQAFVRHMOR 1753
DB 1741 IQLVQAFVRHMOR 1753

RESULT 17
US-10-719-385-19
; Sequence 19, Application US/10719385
; Publication No. US20040209284A1
; GENERAL INFORMATION:
; APPLICANT: O'Toole et al.
; TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis
; FILE REFERENCE: 22058-582
; CURRENT APPLICATION NUMBER: US/10/719,385
; PRIOR FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: PCT/US03/37339
; PRIOR FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: 60/428,094
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 19
; LENGTH: 1753
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-385-19

Query Match 99.8%; Score 8991; DB 17; Length 1753;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1750; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MIRKSKITSVLSFCRSRRLMTLLGRSALRELSQIEABINRWRLLEGLSTYKPPSPS 60
DB 1 MIRKSKITSVLSFCRSRRLMTLLGRSALRELSQIEABINRWRLLEGLSTYKPPSPS 60
QY 61 SAEKVANKOVASPLKELGIRISKFLGLDBEQSVQLQCYLOEDYGRTRDSVTVQDER 120
DB 61 SAEKVANKOVASPLKELGIRISKFLGLDBEQSVQLQCYLOEDYGRTRDSVTVQDER 120
QY 121 QSQALILKTDADYYBERTCILRCVHLILTYFODERHPYREYVADCVDLKEKELVSKYRQ 180
DB 121 QSQALILKTDADYYBERTCILRCVHLILTYFODERHPYREYVADCVDLKEKELVSKYRQ 180
QY 181 FEELYTEAPYMTETHGNLMTERQVSRFVQCLREOSMLLEIFLYAYAFEMASDILLVLT 240
DB 181 FEELYTEAPYMTETHGNLMTERQVSRFVQCLREOSMLLEIFLYAYAFEMASDILLVLT 240
QY 241 KMFKEQFGFSRQTRNRLVDETMDPFVDRIGYSALILVSGMDIESLHKALDDRRELIHQF 300
DB 241 KMFKEQFGFSRQTRNRLVDETMDPFVDRIGYSALILVSGMDIESLHKALDDRRELIHQF 300
QY 301 AODGLICQMDICMLTFGDI PHHAPVLLAMALLRHTLNEBETSSVYRKIGGTAIQLVNQ 360
DB 301 AODGLICQMDICMLTFGDI PHHAPVLLAMALLRHTLNEBETSSVYRKIGGTAIQLVNQ 360
QY 361 YLTRLQSLASGNDCTSTACMCVYGILSPVLTSLHTLNGOODIITPACGVADLPPL 420
DB 361 YLTRLQSLASGNDCTSTACMCVYGILSPVLTSLHTLNGOODIITPACGVADLPPL 420
QY 421 PELFWGTBPTSGILILDSVCGMFPHLSPPLQALRALVSGSTAKVYSLFDKMSFYNE 480
DB 421 PELFWGTBPTSGILILDSVCGMFPHLSPPLQALRALVSGSTAKVYSLFDKMSFYNE 480
QY 481 LYKHKHDVYISHEDGTLMRQTPKLIYPIGGQTNLRIPOGTIVQVNLDRALVYRWEYSY 540
DB 481 LYKHKHDVYISHEDGTLMRQTPKLIYPIGGQTNLRIPOGTIVQVNLDRALVYRWEYSY 540

```



```

Qy 541 SSMFTCEIEMLHVSTADVIQHCORXKPIIDLHVKVISNDLSIADCLPITSRIYML 600
Db 541 SSMFTCEIEMLHVSTADVIQHCORXKPIIDLHVKVISNDLSIADCLPITSRIYML 600
Qy 601 LQRLTVIISPVDVIASCVNCLTVLAARNPAKWTDLRHGTFLPFAHVPVSLSGMISAE 660
Db 601 LQRLTVIISPVDVIASCVNCLTVLAARNPAKWTDLRHGTFLPFAHVPVSLSGMISAE 660
Qy 661 GNNAGGYNLMSNSEQOGEYGVTTAFLRLITTVKGGUGSTQSGVLPCVMFVLEMLP 720
Db 661 GNNAGGYNLMSNSEQOGEYGVTTAFLRLITTVKGGUGSTQSGVLPCVMFVLEMLP 720
Qy 721 SYHKMRVNSHGVREGIGLILIELIHAINTLCHETDLSHSHPSLOFLICISLAITYEAGOT 780
Db 721 SYHKMRVNSHGVREGIGLILIELIHAINTLCHETDLSHSHPSLOFLICISLAITYEAGOT 780
Qy 781 VINIMIGIVDTIDWMAAOPRSDGAGGOGQLIKTVKLAFSVTNNVIRLKPSPNVSP 840
Db 781 VINIMIGIVDTIDWMAAOPRSDGAGGOGQLIKTVKLAFSVTNNVIRLKPSPNVSP 840
Qy 841 LEQALSOHGAGNNILIAVLAKYIYKHDPALPRLAIQLKRLATVAPMSVYACIGNDAAA 900
Db 841 LEQALSOHGAGNNILIAVLAKYIYKHDPALPRLAIQLKRLATVAPMSVYACIGNDAAA 900
Qy 901 IRDAFLRLOSKIEDMRKXWILEFLTVAVETQBLIELFNLKYGQSDSKESKSLGMV 960
Db 901 IRDAFLRLOSKIEDMRKXWILEFLTVAVETQBLIELFNLKYGQSDSKESKSLGMV 960
Qy 961 SCLHVALELISQOQDRYWCPLHRAAIAFLHALMQRDSDAMLVLRTEKPFENLTS 1020
Db 961 SCLHVALELISQOQDRYWCPLHRAAIAFLHALMQRDSDAMLVLRTEKPFENLTS 1020
Qy 1021 LFGTLPSPSESEPSILETCALIMKICLEIYVVVKGSLDOSLKOTLKKSIEKPAWMS 1080
Db 1021 LFGTLPSPSESEPSILETCALIMKICLEIYVVVKGSLDOSLKOTLKKSIEKPAWMS 1080
Qy 1081 GYVSLAHVAETBESSCTSLLEYOMLVASAMRMILITTHADIMHLDVSVRRLPFDV 1140
Db 1081 GYVSLAHVAETBESSCTSLLEYOMLVASAMRMILITTHADIMHLDVSVRRLPFDV 1140
Qy 1141 LDGTRKALLVPASVNCRLGSMKCTLLILRLQMRRELSYDEILIGPTEILEGVLQADQ 1200
Db 1141 LDGTRKALLVPASVNCRLGSMKCTLLILRLQMRRELSYDEILIGPTEILEGVLQADQ 1200
Qy 1201 QLMEXTKAVFSAFITVLQMKEMKVS DIPQYSQVLVANCETLQEEVIALPQTRHSALG 1260
Db 1201 QLMEXTKAVFSAFITVLQMKEMKVS DIPQYSQVLVANCETLQEEVIALPQTRHSALG 1260
Qy 1261 SATEDKOSMETDDGSRSHRQORDGVCTGLHIAELCEVDEDEGSMQVTRRLPITL 1320
Db 1261 SATEDKOSMETDDGSRSHRQORDGVCTGLHIAELCEVDEDEGSMQVTRRLPITL 1320
Qy 1321 LTTLEVSJRMKONLHFTATLHLLTLARTOOGATAVAGAGITGSICTPLSVYOLSTNG 1380
Db 1321 LTTLEVSJRMKONLHFTATLHLLTLARTOOGATAVAGAGITGSICTPLSVYOLSTNG 1380
Qy 1381 TAQTPSASRKSILDASWPGVYRLSMSLMEQLLKTLYNLFLEPALDFVGHQERTLOCLNA 1440
Db 1381 TAQTPSASRKSILDASWPGVYRLSMSLMEQLLKTLYNLFLEPALDFVGHQERTLOCLNA 1440
Qy 1441 VRTVQSLACLEBADHTVGFILQLSNFMKEMHFLPOLMRDIOVNLGYICOACTSLSHSRK 1500
Db 1441 VRTVQSLACLEBADHTVGFILQLSNFMKEMHFLPOLMRDIOVNLGYICOACTSLSHSRK 1500
Qy 1501 MLQHLQKNGDGLPSAVALQVQRPSPASAPSSSKOPADTEASEBOALHTVOYGLK 1560
Db 1501 MLQHLQKNGDGLPSAVALQVQRPSPASAPSSSKOPADTEASEBOALHTVOYGLK 1560
Qy 1561 ILSTKLALRHFTPDVQIILLDOSLIDLABYNFLFALSFTTPTDSEVAPSPGTLIATVNV 1620
Db 1561 ILSTKLALRHFTPDVQIILLDOSLIDLABYNFLFALSFTTPTDSEVAPSPGTLIATVNV 1620

```

```

Qy 1621 ALNMLGELDKKKEPLTOAVGSTOAGETRLKSLMFTMENC FYLLISQAMRYLDRPAVH 1680
Db 1621 ALNMLGELDKKKEPLTOAVGSTOAGETRLKSLMFTMENC FYLLISQAMRYLDRPAVH 1680
Qy 1681 PRDQKMKQELISELSTLSSLSRYFRRGAPSPATGYLPSPQGSKTSLSQASPESSQPL 1740
Db 1681 PRDQKMKQELISELSTLSSLSRYFRRGAPSPATGYLPSPQGSKTSLSQASPESSQPL 1740
Qy 1741 IOLVOAFVRHMOR 1753
Db 1741 IOLVOAFVRHMOR 1753

RESULT 18
US-10-719-385-5
; Sequence 5, Application US/10719385
; Publication No. US20040209284A1
; GENERAL INFORMATION:
; APPLICANT: O'Toole et al.
; TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis
; FILE REFERENCE: 22058-582
; CURRENT APPLICATION NUMBER: US/10/719,385
; PRIOR FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: PCT/US03/37339
; PRIOR FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: 60/428,094
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 5
; LENGTH: 1752
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-385-5

Query Match 99.8%; Score 8985.5; DB 17; Length 1752;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1751; Conservative 1; Mismatches 0; Indels 1; Gaps 1;

Qy 1 MIRSKITSVTSFCRSSRELTITLLGRSALBELSOIEALNKMRRLLEGISTYKPPSPS 60
Db 1 MIRSKITSVTSFCRSSRELTITLLGRSALBELSOIEALNKMRRLLEGISTYKPPSPS 60
Qy 61 SAEKVRANKDVASPLKEIGLRISKPLGIDEBOSVOLQCYOEDRGTRDSKYTLQDER 120
Db 61 SAEKVRANKDVASPLKEIGLRISKPLGIDEBOSVOLQCYOEDRGTRDSKYTLQDER 120
Qy 121 OSQALILKIADYYEERTCILRCVHLHLYTFQDERHPYRVEYADCVDKLEKELVSKYRQ 180
Db 121 OSQALILKIADYYEERTCILRCVHLHLYTFQDERHPYRVEYADCVDKLEKELVSKYRQ 180
Qy 181 FEEIYKTEAPWETHGNLINTERQVSRWFVQCLREOSMLLEIIPLYYAVFENAPSDLVLT 240
Db 181 FEEIYKTEAPWETHGNLINTERQVSRWFVQCLREOSMLLEIIPLYYAVFENAPSDLVLT 240
Qy 241 KMFKEQGFSGQTRNHLVDETMDFVDKIGYFSALIVEGNDIBSLHRCALDDREELHOF 300
Db 241 KMFKEQGFSGQTRNHLVDETMDFVDKIGYFSALIVEGNDIBSLHRCALDDREELHOF 300
Qy 301 AODGLICQDMQCLMLTFGDI PHHAPVLLAMALLRHTLNPEBTSYVRKIGGTALQANFQ 360
Db 301 AODGLICQDMQCLMLTFGDI PHHAPVLLAMALLRHTLNPEBTSYVRKIGGTALQANFQ 360
Qy 361 YLTRLQSLAGSAGNDCTTSTACMCVYGLSFLVTSLEHLTGNODIIDTACEVLADPSL 420
Db 361 YLTRLQSLAGSAGNDCTTSTACMCVYGLSFLVTSLEHLTGNODIIDTACEVLADPSL 420
Qy 421 PELFWGTEPTSGILGILDSVCGMPHLSPLLOLRLALVSGSTAKKYSLDKKSFNE 480
Db 421 PELFWGTEPTSGILGILDSVCGMPHLSPLLOLRLALVSGSTAKKYSLDKKSFNE 480
Qy 481 LYKHKPHDVISHEDGTLRRQTPKLLYPLGGQTNLRIPQGTIVGQVMDRAVLYRWEYSY 540
Db 481 LYKHKPHDVISHEDGTLRRQTPKLLYPLGGQTNLRIPQGTIVGQVMDRAVLYRWEYSY 540

```

480 LYKHHPHVISHEDGTLWRROTPKLLYPLGGOITNLRIPQGTGVQVMDLDRAYLVWREYSY 539  
QY SSWTLFTEIEMMLHVVSTADVIQHCQVKKPIIDLVHNVISTDLSADCLPITSRIYML 600  
Db SSWTLFTEIEMMLHVVSTADVIQHCQVKKPIIDLVHNVISTDLSADCLPITSRIYML 599  
QY LORLTVISPPVDVIASCVNCLTVLAANPAKVTDLNHTGFLPVFAHVSLSQMSIAE 660  
Db LORLTVISPPVDVIASCVNCLTVLAANPAKVTDLNHTGFLPVFAHVSLSQMSIAE 659  
QY GMAAGYGNLNMNSBQPOGEYGVTTIAPLRLITTVKQGLSTQSGVLPVCMFVKEMLP 720  
Db GMAAGYGNLNMNSBQPOGEYGVTTIAPLRLITTVKQGLSTQSGVLPVCMFVKEMLP 719  
QY SYHKRVNHSQVREIGICLILIELHAIINLCHETDHSHPISLOFLICSLATEAGOT 780  
Db SYHKRVNHSQVREIGICLILIELHAIINLCHETDHSHPISLOFLICSLATEAGOT 779  
QY VININGIGVDTIDWMAAQRSDGAGGQGLIKITVKLAFSVTNNVIRLKPSNVVSP 840  
Db VININGIGVDTIDWMAAQRSDGAGGQGLIKITVKLAFSVTNNVIRLKPSNVVSP 839  
QY LEQALSOHAGHNNIIAVLAKYIYKHPDLPRLAIQLKRLATVAPMSVYACIGNDAAA 900  
Db LEQALSOHAGHNNIIAVLAKYIYKHPDLPRLAIQLKRLATVAPMSVYACIGNDAAA 899  
QY IRDALFTLQSKIEMRIKMWLEFLTVAVETOPBLIFLNLKXKSDSKESLGMW 960  
Db IRDALFTLQSKIEMRIKMWLEFLTVAVETOPBLIFLNLKXKSDSKESLGMW 959  
QY IRDALFTLQSKIEMRIKMWLEFLTVAVETOPBLIFLNLKXKSDSKESLGMW 959  
Db IRDALFTLQSKIEMRIKMWLEFLTVAVETOPBLIFLNLKXKSDSKESLGMW 959  
QY SCIAHVLIELDSQODRYWCPRLHRAAIAFLHALMODRDSAMLVKTKRFENMLTSP 1020  
Db SCIAHVLIELDSQODRYWCPRLHRAAIAFLHALMODRDSAMLVKTKRFENMLTSP 1019  
QY LFTLSPSESTSEPSILETCALIMKICLEIYVYVKSGLDSGLKOTLKKSIEKPAWVS 1080  
Db LFTLSPSESTSEPSILETCALIMKICLEIYVYVKSGLDSGLKOTLKKSIEKPAWVS 1079  
QY GYVKSIAHVAVATESSCTSLLEYOMLVSAMRMLIINTHADINHLDSVVRQLPFDV 1140  
Db GYVKSIAHVAVATESSCTSLLEYOMLVSAMRMLIINTHADINHLDSVVRQLPFDV 1139  
QY LDGTALLLVPASVAVCLRLGSMKCTLLILLRQMKRELSVDEILGPTLEILEGLQADQ 1200  
Db LDGTALLLVPASVAVCLRLGSMKCTLLILLRQMKRELSVDEILGPTLEILEGLQADQ 1199  
QY QIMERTKAKVPSAFITTVLQMKEMKVSDFPOYSQVLVNCETLQEBVIALFDOTRHSIALG 1259  
Db QIMERTKAKVPSAFITTVLQMKEMKVSDFPOYSQVLVNCETLQEBVIALFDOTRHSIALG 1259  
QY SATTEKDSWETDCCRSRHRDQDCVYLGHLAKELCEVDEGDGSMLOVTRRLPILPTL 1320  
Db SATTEKDSWETDCCRSRHRDQDCVYLGHLAKELCEVDEGDGSMLOVTRRLPILPTL 1319  
QY LTTLEVSILRMKONLFTTEATLHLLTLARTOOGATAVAGAGITOSICLPLLSVYOLSTNG 1380  
Db LTTLEVSILRMKONLFTTEATLHLLTLARTOOGATAVAGAGITOSICLPLLSVYOLSTNG 1379  
QY LTTLEVSILRMKONLFTTEATLHLLTLARTOOGATAVAGAGITOSICLPLLSVYOLSTNG 1379  
Db LTTLEVSILRMKONLFTTEATLHLLTLARTOOGATAVAGAGITOSICLPLLSVYOLSTNG 1379  
QY TAQTPSASRKSIDAPSPGVYRLSWSLMEQLIKTIRYNFLPEALDFVGHQERTLQICANA 1439  
Db TAQTPSASRKSIDAPSPGVYRLSWSLMEQLIKTIRYNFLPEALDFVGHQERTLQICANA 1439  
QY VRTVOSIACLEADHVGFTLQLSNPMKEMHFLPOLMMDIQVNLGYLCOACTSLHSRK 1500  
Db VRTVOSIACLEADHVGFTLQLSNPMKEMHFLPOLMMDIQVNLGYLCOACTSLHSRK 1499  
QY VRTVOSIACLEADHVGFTLQLSNPMKEMHFLPOLMMDIQVNLGYLCOACTSLHSRK 1499  
Db VRTVOSIACLEADHVGFTLQLSNPMKEMHFLPOLMMDIQVNLGYLCOACTSLHSRK 1499  
QY MLOHTLQNKQDGLSVAQVORVOPSPASASAPSSSKQPADTEASEQOALHTVOYGLX 1560  
Db MLOHTLQNKQDGLSVAQVORVOPSPASASAPSSSKQPADTEASEQOALHTVOYGLX 1559  
QY MLOHTLQNKQDGLSVAQVORVOPSPASASAPSSSKQPADTEASEQOALHTVOYGLX 1559  
Db MLOHTLQNKQDGLSVAQVORVOPSPASASAPSSSKQPADTEASEQOALHTVOYGLX 1559  
QY ILSKTLAALRHFTTPDYCOILLDOSIDLAEYNFLFALSFTPTPDSVAPSGTLLATVAV 1620  
Db ILSKTLAALRHFTTPDYCOILLDOSIDLAEYNFLFALSFTPTPDSVAPSGTLLATVAV 1619

QY 1621 ALNMLGELDKKKEPLTQAVGLSTQABGTRTLKSLMFTMENCYLLISQAMRYLRDPAVH 1680  
Db 1620 ALNMLGELDKKKEPLTQAVGLSTQABGTRTLKSLMFTMENCYLLISQAMRYLRDPAVH 1679  
QY 1681 PRDKORMKOEISSELSTLSSLSRYPFRGAPSSPATCVLPSPQKSTSLSKASPESEGPL 1740  
Db 1680 PRDKORMKOEISSELSTLSSLSRYPFRGAPSSPATCVLPSPQKSTSLSKASPESEGPL 1739  
QY 1741 IOLVOAFVRHMOR 1753  
Db 1740 IOLVOAFVRHMOR 1752

RESULT 19  
US-10-719-385-21  
; Sequence 21, Application US/10719385  
; Publication No. US20040209284A1  
; GENERAL INFORMATION:  
; APPLICANT: O'Toole et al.  
; TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis  
; FILE REFERENCE: 22058-582  
; CURRENT APPLICATION NUMBER: US/10/719,385  
; PRIOR FILING DATE: 2003-11-21  
; PRIOR APPLICATION NUMBER: PCT/US03/37339  
; PRIOR FILING DATE: 2003-11-21  
; PRIOR APPLICATION NUMBER: 60/428,094  
; NUMBER OF SEQ ID NOS: 26  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 21  
; LENGTH: 1745  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-719-385-21

Query Match 99.2%; Score 8939; DB 17; Length 1745;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1739; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 15 RSSREMLTILGRALRELSQIEAEINMKWRRLLEGISYKKPPSSAEKYKANKDVASP 74  
Db 7 RSSREMLTILGRALRELSQIEAEINMKWRRLLEGISYKKPPSSAEKYKANKDVASP 66  
QY LKEIGLRIKSKPLGDECSQVQLQCYLQEDYRGTRDSVKTVLQDEROSQALILKXIADYYY 134  
Db LKEIGLRIKSKPLGDECSQVQLQCYLQEDYRGTRDSVKTVLQDEROSQALILKXIADYYY 126  
QY 135 EERTCILRCVLIHLTYFQDERHPYRVEYADCVDKLEKELVSKYRQOFEEELYKTEAPTWET 194  
Db 127 EERTCILRCVLIHLTYFQDERHPYRVEYADCVDKLEKELVSKYRQOFEEELYKTEAPTWET 186  
QY 195 HGNLMTROVSRMNVQCLARESMLEETIPLYAAFEWAPSLVLTQMKFGSGSGRQTN 254  
Db 187 HGNLMTROVSRMNVQCLARESMLEETIPLYAAFEWAPSLVLTQMKFGSGSGRQTN 246  
QY 255 RHLVDETMDPEVDRDIGFSAIILVEGMDIESLHKCALDRRELHQAQDGLICQDMDCLM 314  
Db 247 RHLVDETMDPEVDRDIGFSAIILVEGMDIESLHKCALDRRELHQAQDGLICQDMDCLM 306  
QY 315 LTFGDIPIHHAVALAMALLRHTLNPBETSSVYRKIGGTALQANFOYLTRLOSLASGSGN 374  
Db 307 LTFGDIPIHHAVALAMALLRHTLNPBETSSVYRKIGGTALQANFOYLTRLOSLASGSGN 366  
QY 375 DCTSTACMCYGLSLPULSTLEHTLGNODIIDTACEVADPSLPFLFGTEPTSGIG 434  
Db 367 DCTSTACMCYGLSLPULSTLEHTLGNODIIDTACEVADPSLPFLFGTEPTSGIG 426  
QY 435 IILDSVCGMFPHLSPILLQALRALVSGKSTAKKYVSFLDKMSFYNELYKHKPHDVI SHED 494  
Db 427 IILDSVCGMFPHLSPILLQALRALVSGKSTAKKYVSFLDKMSFYNELYKHKPHDVI SHED 486  
QY 495 GTLWRROTPKLLYPLGGOITNLRIPQGTGVQVMDLDRAYLVWREYSYSSWTLFTEIEMML 554

```

Db 487 GTWRGQPKLLPLPGSGOTNLRPGTGVQWLDDBRALVLRNREYSSTWLTCTCIEML 546
Qy 555 HVSSTADVIQHCQVRKPIIDLVHKVISTDLSTADCLPITSRIYMLQRLTTVISPVDV 614
Db 547 HVSSTADVIQHCQVRKPIIDLVHKVISTDLSTADCLPITSRIYMLQRLTTVISPVDV 606
Qy 615 IASCNCLTVLAARPAKVTMDLRHTGFLPVAHVSSLSQMTSAGGNAGGYGMLNNS 674
Db 607 IASCNCLTVLAARPAKVTMDLRHTGFLPVAHVSSLSQMTSAGGNAGGYGMLNNS 666
Qy 675 EPOGEYVTTIAFLRLITTLVKQSGSTOSQGLVPCWVFLKEMLPYHKRKNYSHGYRE 734
Db 667 EPOGEYVTTIAFLRLITTLVKQSGSTOSQGLVPCWVFLKEMLPYHKRKNYSHGYRE 726
Qy 735 QIGCLILHAIHAIHMLCHETDLHSHSTPSLOFLCISLAYTEAGGTINIMIGVDTIDM 794
Db 727 QIGCLILHAIHAIHMLCHETDLHSHSTPSLOFLCISLAYTEAGGTINIMIGVDTIDM 786
Qy 795 VMAAPRSDGAGGOGQGLIKTVLAFSYNNVRLRPPSNVPSLEOALSOHAGNN 854
Db 787 VMAAPRSDGAGGOGQGLIKTVLAFSYNNVRLRPPSNVPSLEOALSOHAGNN 846
Qy 855 LIAVLAKYIYKHDPALRLAIQLKRLATVAPMSVYACLGNDAAIRDAFLTRLOSKE 914
Db 847 LIAVLAKYIYKHDPALRLAIQLKRLATVAPMSVYACLGNDAAIRDAFLTRLOSKE 906
Qy 915 DMRIKVMLEPLTVAVETOPGLIELFLNLEVKDSDGSKESP9LGWMSCLHAYLELIDSQ 974
Db 907 DMRIKVMLEPLTVAVETOPGLIELFLNLEVKDSDGSKESP9LGWMSCLHAYLELIDSQ 966
Qy 975 ODRWCPPLHRAALAFILHMLMODRDSAMVLRKPKFWMLTSPLEGTLSPESTSEP 1034
Db 967 ODRWCPPLHRAALAFILHMLMODRDSAMVLRKPKFWMLTSPLEGTLSPESTSEP 1026
Qy 1035 SILETCALIMKIICLEIYVYVKSGLDOSLKTLLKFFSIEKRAYSGYKSLAVNASTE 1094
Db 1027 SILETCALIMKIICLEIYVYVKSGLDOSLKTLLKFFSIEKRAYSGYKSLAVNASTE 1086
Qy 1095 GSSCTSLLEYOMVSAMWMLIATTHADIMHLDVSVRROIFLVDLDTKALLVPAV 1154
Db 1087 GSSCTSLLEYOMVSAMWMLIATTHADIMHLDVSVRROIFLVDLDTKALLVPAV 1146
Qy 1155 NCLRIGSKKCTLLILRQMKRELGSVDEILGPIELIEGVQADQOQMEKTKAVFSAF 1214
Db 1147 NCLRIGSKKCTLLILRQMKRELGSVDEILGPIELIEGVQADQOQMEKTKAVFSAF 1206
Qy 1215 ITVLQMKEMKYSIDIPQYSOLVNVCELTQOEBVIALFDOTRSHLAGSATBEDKSDMETDC 1274
Db 1207 ITVLQMKEMKYSIDIPQYSOLVNVCELTQOEBVIALFDOTRSHLAGSATBEDKSDMETDC 1266
Qy 1275 SRSRRDRDQVCVGLHAKELCEVDEBDGSMLOVTRRLPILPTLLTTLVSLMKONL 1334
Db 1267 SRSRRDRDQVCVGLHAKELCEVDEBDGSMLOVTRRLPILPTLLTTLVSLMKONL 1326
Qy 1335 HFTXTTLLHLLATQOGATVAVAGITQSCLEPLSVYQULSTNGTQOTPBASKSIDA 1394
Db 1327 HFTXTTLLHLLATQOGATVAVAGITQSCLEPLSVYQULSTNGTQOTPBASKSIDA 1386
Qy 1395 PSMPEVYRLSLSMLQKLTLRYNFLPEALDFVGHOBERTLOCLNAVTVOSIACLEBAD 1454
Db 1387 PSMPEVYRLSLSMLQKLTLRYNFLPEALDFVGHOBERTLOCLNAVTVOSIACLEBAD 1446
Qy 1455 HTVGFILOLSNFMKWHFHLPOLMDIOVNIAGYLCOACTSLHSHKMLQHYLQNRNGDL 1514
Db 1447 HTVGFILOLSNFMKWHFHLPOLMDIOVNIAGYLCOACTSLHSHKMLQHYLQNRNGDL 1506
Qy 1515 PSAAVQVRQPPSASAPSSSKOPADTEASEOQALHTVOYGLKILSKITLAARHPTP 1574
Db 1507 PSAAVQVRQPPSASAPSSSKOPADTEASEOQALHTVOYGLKILSKITLAARHPTP 1566
Qy 1575 DVCCQLLDQSLDLAEYNFLFALSFTTPTFDSVAPSGTILATVNAVLMGELDKKEP 1634

```

```

Db 1567 DVCCQLLDQSLDLAEYNFLFALSFTTPTFDSVAPSGTILATVNAVLMGELDKKEP 1626
Qy 1635 LTOAVGLSTQAEGRITLKSLLMFTMENCYLLISQANRYLDDPAVHPDKQRMQELSS 1694
Db 1627 LTOAVGLSTQAEGRITLKSLLMFTMENCYLLISQANRYLDDPAVHPDKQRMQELSS 1686
Qy 1695 LSTLLSLSRFRGAPSPATGVLPSPQGSTLSKSPSPOEPLIOVQAFVHMOR 1753
Db 1687 LSTLLSLSRFRGAPSPATGVLPSPQGSTLSKSPSPOEPLIOVQAFVHMOR 1745

RESULT 20
US-10-370-715B-544
; Sequence 544, Application US/10370715B
; Publication No. US20040258678A1
; GENERAL INFORMATION:
; Patin Docket Preview
; APPLICANT: BODARY, SARAH C.
; APPLICANT: CLARK, HILARY
; APPLICANT: BRISDELL, HUNTER
; APPLICANT: JACKMAN, JARRET
; APPLICANT: SCHOENFELD, JILL R.
; APPLICANT: WILLIAMS, P. MICKEY
; APPLICANT: WOOD, WILLIAM I.
; APPLICANT: MU, THOMAS D.
; TITLE OF INVENTION: Compositions and Methods for the Treatment of Immune
; FILE REFERENCE: P1948R1-US
; CURRENT FILING DATE: 2003-02-21
; NUMBER OF SEQ ID NOS: 742
; SEQ ID NO 544
; LENGTH: 1745
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-370-715B-544

Query Match 99.2%; Score 8939; DB 17; Length 1745;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1739; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 15 RSSRELWTLLGRSALRELSQIEAELNGMRRLLEGSLYKPPSPSSAEKYANKDVASP 74
Db 7 RSSRELWTLLGRSALRELSQIEAELNGMRRLLEGSLYKPPSPSSAEKYANKDVASP 66
Qy 75 LKEGLRISKFLGLDEBQSVLLQCYIQEDYRGRDSVKTVALQDERQOALILKTADYYY 134
Db 67 LKEGLRISKFLGLDEBQSVLLQCYIQEDYRGRDSVKTVALQDERQOALILKTADYYY 126
Qy 135 EERTCILACVHLTLTFODERHPRAVEADCVDLKLEKLVSKYQOFEELYKTAPPTMET 194
Db 127 EERTCILACVHLTLTFODERHPRAVEADCVDLKLEKLVSKYQOFEELYKTAPPTMET 186
Qy 195 HGNLMTQVRWRFQOCIREOSMLLEIFLYAAYFEMAPSLVLTCKAFKEQSGFSRQTN 254
Db 187 HGNLMTQVRWRFQOCIREOSMLLEIFLYAAYFEMAPSLVLTCKAFKEQSGFSRQTN 246
Qy 255 RHLVDETMDPVDRIGYFSALILVEGMDIESLHKCALDRRELHQFADGILCQMDCLM 314
Db 247 RHLVDETMDPVDRIGYFSALILVEGMDIESLHKCALDRRELHQFADGILCQMDCLM 306
Qy 315 LTFQGIPIHRAVLLAMALLRHTLNDPEETSSVVRKIGTAIQLVNFQVYLRLLQSLASGN 374
Db 307 LTFQGIPIHRAVLLAMALLRHTLNDPEETSSVVRKIGTAIQLVNFQVYLRLLQSLASGN 366
Qy 375 DCTSTACMCYGLLSFVLTSLLEHTTGNQODIDTACEVLADPSLPFLFWGTBPTSGLG 434
Db 367 DCTSTACMCYGLLSFVLTSLLEHTTGNQODIDTACEVLADPSLPFLFWGTBPTSGLG 426
Qy 435 IILDSVCMFPHLLSPILLQILRALVSGSKTAKKYVSFLDKXSFYNELYKHPHDVISHED 494
Db 427 IILDSVCMFPHLLSPILLQILRALVSGSKTAKKYVSFLDKXSFYNELYKHPHDVISHED 486

```

QY 495 GTLMRRQTPKLLYPLGGQTNLRIPOGTVGQVWLDDBRAVLVRWEYSYSSWTLFTCEIEMLL 554  
 DB 487 GTLMRRQTPKLLYPLGGQTNLRIPOGTVGQVWLDDBRAVLVRWEYSYSSWTLFTCEIEMLL 546  
 QY 555 HVVSTADYVQHOCORVYPIIDLVHKVISTDLSTADCLLPITSRIYMLQRLTTVISPPVDV 614  
 DB 547 HVVSTADYVQHOCORVYPIIDLVHKVISTDLSTADCLLPITSRIYMLQRLTTVISPPVDV 606  
 QY 615 IASCNCLTVLAARPAKVTMDLRHTGFLPFVAHPVSSLSOMISAEGNAGGYGLMLNS 674  
 DB 607 IASCNCLTVLAARPAKVTMDLRHTGFLPFVAHPVSSLSOMISAEGNAGGYGLMLNS 666  
 QY 675 EOPQGEYGVTAFLRLITTLVKQSGSQGLVPCVAFVLEKMLPSYHKMRKYNHSGVRE 734  
 DB 667 EOPQGEYGVTAFLRLITTLVKQSGSQGLVPCVAFVLEKMLPSYHKMRKYNHSGVRE 726  
 QY 735 QIGCIIIBELIHAIMLCHETDLSHSHTSLOPLCISLAYTBAGOTVINIGIVDTIDM 794  
 DB 727 QIGCIIIBELIHAIMLCHETDLSHSHTSLOPLCISLAYTBAGOTVINIGIVDTIDM 786  
 QY 795 VMAAPRSDGAEQSGQGLIKTVLAFTNNVIRLPPSNVVSPLRQALSOHAGHGN 854  
 DB 787 VMAAPRSDGAEQSGQGLIKTVLAFTNNVIRLPPSNVVSPLRQALSOHAGHGN 846  
 QY 855 LIAVLAKYIYHKGDPALPRLAIQLKRLATVAPMSVYACLGNDAAIRDAEFLTRLOSKE 914  
 DB 847 LIAVLAKYIYHKGDPALPRLAIQLKRLATVAPMSVYACLGNDAAIRDAEFLTRLOSKE 906  
 QY 915 DMRIKVMLEFLTVAVETOPGLIELFNLVYKDGSGSKESLSGWSCLAHVBLIDSQ 974  
 DB 907 DMRIKVMLEFLTVAVETOPGLIELFNLVYKDGSGSKESLSGWSCLAHVBLIDSQ 966  
 QY 975 QDRYWCPLPRLRAAIAFLHALMODRDSAMLVLRTPKPFWMNLSPLEFGLTSPSETSEP 1034  
 DB 967 QDRYWCPLPRLRAAIAFLHALMODRDSAMLVLRTPKPFWMNLSPLEFGLTSPSETSEP 1026  
 QY 1035 SILETCALIMKIICLEIYVYVKSIDQSLKDTLKKSIEKRPAYWSGYKSLAHVAETE 1094  
 DB 1027 SILETCALIMKIICLEIYVYVKSIDQSLKDTLKKSIEKRPAYWSGYKSLAHVAETE 1086  
 QY 1095 GSSCSTILEYOMLVSAMRMLIATTHADIMHLDVSVRROLFDVLDGTALLVPAV 1154  
 DB 1087 GSSCSTILEYOMLVSAMRMLIATTHADIMHLDVSVRROLFDVLDGTALLVPAV 1146  
 QY 1155 NCLRSGNCKTLLILRLQMKRELGSVDEILGPLETEILEGYLOADQOLMEKTKAVSAF 1214  
 DB 1147 NCLRSGNCKTLLILRLQMKRELGSVDEILGPLETEILEGYLOADQOLMEKTKAVSAF 1206  
 QY 1215 ITVYQMKEMKYSDFQYSQVLVANCETIQOEVIYALFDOTRHSIALGATBEDXDMETDDC 1274  
 DB 1207 ITVYQMKEMKYSDFQYSQVLVANCETIQOEVIYALFDOTRHSIALGATBEDXDMETDDC 1266  
 QY 1275 SRSRHRDQDGVYGLHLAKELCEVDEDDGSMLOVTRRLPILPTLITTEVSLRMKNL 1334  
 DB 1267 SRSRHRDQDGVYGLHLAKELCEVDEDDGSMLOVTRRLPILPTLITTEVSLRMKNL 1326  
 QY 1335 HFEATLHLLITLARQOGATVAVAGITQISICPLSVYQSLSTNGTAQTPSASRKSIDA 1394  
 DB 1327 HFEATLHLLITLARQOGATVAVAGITQISICPLSVYQSLSTNGTAQTPSASRKSIDA 1386  
 QY 1395 PSWNGVYRLSWSLMEQLIKTRYNFLPEALDFVGHQERTLOCNNAVTVOSLACEAD 1454  
 DB 1387 PSWNGVYRLSWSLMEQLIKTRYNFLPEALDFVGHQERTLOCNNAVTVOSLACEAD 1446  
 QY 1455 HTVEFIILQSLNFKEMHFLPOLMRDIOVNLGYLCOACTSLHSRKMLOHLOKNDG 1514  
 DB 1447 HTVEFIILQSLNFKEMHFLPOLMRDIOVNLGYLCOACTSLHSRKMLOHLOKNDG 1506  
 QY 1515 PSAVAQVQRPSPASASAPSSSKOPADTEASEOALHTVQYGLIKLITKIALRHF 1574  
 DB 1507 PSAVAQVQRPSPASASAPSSSKOPADTEASEOALHTVQYGLIKLITKIALRHF 1566  
 QY 1575 DVQOILDDQSLDLAEVNFPLALSTFTTPDSEVAPSGEITLAVNALMNLGELDKKEP 1634

DB 1567 DVQOILDDQSLDLAEVNFPLALSTFTTPDSEVAPSGEITLAVNALMNLGELDKKEP 1626  
 QY 1635 LTOAVGLSTQAEGRITLKSILMTWENCYFYLISQAMRYLRDPAVHRDQRMKQELSS 1694  
 DB 1627 LTOAVGLSTQAEGRITLKSILMTWENCYFYLISQAMRYLRDPAVHRDQRMKQELSS 1686  
 QY 1695 LSTLSSLSRYFRGASPPATGVLPSPOKSTLSASPSOEPLOVOAPFRHQR 1753  
 DB 1687 LSTLSSLSRYFRGASPPATGVLPSPOKSTLSASPSOEPLOVOAPFRHQR 1745  
 RESULT 21  
 US-10-719-385-22  
 ; Sequence 22, Application US/10719385  
 ; Publication No. US20040209284A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: O'Toole et al.  
 ; TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis  
 ; FILE REFERENCE: 22058-582  
 ; CURRENT APPLICATION NUMBER: US/10/719,385  
 ; PRIOR FILING DATE: 2003-11-21  
 ; PRIOR APPLICATION NUMBER: PCT/US03/37339  
 ; PRIOR FILING DATE: 2003-11-21  
 ; NUMBER OF SEQ ID NOS: 26  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 22  
 ; LENGTH: 3534  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-10-719-385-22  
 Query Match 99.1%; Score 8927; DB 17; Length 3534;  
 Best Local Similarity 99.8%; Pred. No. 0;  
 Matches 1739; Conservative 0; Mismatches 0; Indels 4; Gaps 1;  
 QY 15 RSSRELWTLIGRSALRELSQIEAEINKWRRLLEGLSYKPPSPSAEKYKANKDVASP 74  
 DB 25 RSSRELWTLIGRSALRELSQIEAEINKWRRLLEGLSYKPPSPSAEKYKANKDVASP 84  
 QY 75 LKEIGLRISKFLGLDEQSVOLLQCYLOEDYRGTDSVKTVALODERQOALLIKADYVY 134  
 DB 85 LKEIGLRISKFLGLDEQSVOLLQCYLOEDYRGTDSVKTVALODERQOALLIKADYVY 144  
 QY 135 EERTCLICVHLITTYFQDERHRYRYEVADCVKLEKELVSKYRQOFEBELYKTEAPTWET 194  
 DB 145 EERTCLICVHLITTYFQDERHRYRYEVADCVKLEKELVSKYRQOFEBELYKTEAPTWET 204  
 QY 195 HGNLMTERRQVSRWFVQCLREOSMLLEIIPLYAYFEMAPSDLLVLRKMFQSGFSGRQTN 254  
 DB 205 HGNLMTERRQVSRWFVQCLREOSMLLEIIPLYAYFEMAPSDLLVLRKMFQSGFSGRQTN 264  
 QY 255 RHLVDETMDPFVDRIGYFSALIIIVEGMDIESLHKCALDDBRELHQAODGLICQMDCLM 314  
 DB 265 RHLVDETMDPFVDRIGYFSALIIIVEGMDIESLHKCALDDBRELHQAODGLICQMDCLM 324  
 QY 315 LTRGDIPIHAPVLLAVALRHTLNPEETSIVYKIGTGAIQLVNROYLRILLOSLSGSGN 374  
 DB 325 LTRGDIPIHAPVLLAVALRHTLNPEETSIVYKIGTGAIQLVNROYLRILLOSLSGSGN 384  
 QY 375 DCTTSTACMCVYGLSFVLTSLAFLTGNOODIIPRACEVADPSLPFLFMGTEPTSGIG 434  
 DB 385 DCTTSTACMCVYGLSFVLTSLAFLTGNOODIIPRACEVADPSLPFLFMGTEPTSGIG 444  
 QY 435 IILDVCGMFPHLISPLQLLRALVSGKSTAKKVYSFLDKMSPFYNELYKHKPHDIVSHED 494  
 DB 445 IILDVCGMFPHLISPLQLLRALVSGKSTAKKVYSFLDKMSPFYNELYKHKPHDIVSHED 504  
 QY 495 GTLMRRQTPKLLYPLGGQTNLRIPOGTVGQVWLDDBRAVLVRWEYSYSSWTLFTCEIEMLL 554  
 DB 505 GTLMRRQTPKLLYPLGGQTNLRIPOGTVGQVWLDDBRAVLVRWEYSYSSWTLFTCEIEMLL 564

```

Qy 555 HVSTADVIQHCORVKPIIDLVHKVISTDLADIADCLPITSRIYMLQRLTTVISPPDV 614
Db 565 HVSSTADVIQHCORVKPIIDLVHKVISTDLADIADCLPITSRIYMLQRLTTVISPPDV 624
Qy 615 IASCVNCLTVLAARNPAKWTDLRHGTGFLPFVAHVSSLSQMSAEGNAGGYNLMS 674
Db 625 IASCVNCLTVLAARNPAKWTDLRHGTGFLPFVAHVSSLSQMSAEGNAGGYNLMS 684
Qy 675 EQPGEYGVITAFRLITTLVKQGLSTQSGQLVPCWVFLKEMLPSTHKRRYNSHGVR 734
Db 685 EQPGEYGVITAFRLITTLVKQGLSTQSGQLVPCWVFLKEMLPSTHKRRYNSHGVR 744
Qy 735 QIGCLIELHAHNLCHETDLSHSTPSLOPLCISLAEYBAGQTVNINIGVDTDM 794
Db 745 QIGCLIELHAHNLCHETDLSHSTPSLOPLCISLAEYBAGQTVNINIGVDTDM 804
Qy 795 VMAAPSDGAEQGGQGLIKTVLAFTSVNNVIRLKPNSNVSPLEQALSHGAGNN 854
Db 805 VMAAPSDGAEQGGQGLIKTVLAFTSVNNVIRLKPNSNVSPLEQALSHGAGNN 864
Qy 855 LIATLAKTIYKHDPAFLRAIOLIKRLATVAPMSVACLGNDAAIRDAFLTRLOSKE 914
Db 865 LIATLAKTIYKHDPAFLRAIOLIKRLATVAPMSVACLGNDAAIRDAFLTRLOSKE 924
Qy 915 DMRKVMLEFLTAVETOPGLIELFLNLVYKDSGSKESLGWMSGLHVALELDSQ 974
Db 925 DMRKVMLEFLTAVETOPGLIELFLNLVYKDSGSKESLGWMSGLHVALELDSQ 984
Qy 975 QDRYWCPEPLHRAAIAFLHALMODRDSAMLVLRTPKFWENLISPLFGTSPSETSEP 1034
Db 985 QDRYWCPEPLHRAAIAFLHALMODRDSAMLVLRTPKFWENLISPLFGTSPSETSEP 1044
Qy 1035 SILETCALIMKICIEIYVVVKSIDQSLKDTLKKFSIEKRPAYSGVYKSLAVHAETE 1094
Db 1045 SILETCALIMKICIEIYVVVKSIDQSLKDTLKKFSIEKRPAYSGVYKSLAVHAETE 1104
Qy 1095 GSSCSTSLLEYOMLVASAMRMLIATTHADIMHLDVSVRRLOFLDVLGTALLLVPSV 1154
Db 1105 GSSCSTSLLEYOMLVASAMRMLIATTHADIMHLDVSVRRLOFLDVLGTALLLVPSV 1164
Qy 1155 NCLRISGNKCTLLILLRQMK-----RELGSYDEILGPLEILBGLVQADQOLMEKTKAV 1210
Db 1165 NCLRISGNKCTLLILLRQMK-----RELGSYDEILGPLEILBGLVQADQOLMEKTKAV 1224
Qy 1211 FSAFITYLOMKEMKVSIDIPQYSQVLVANCETLOEFVIALPQTRHSIALGSATEDKDSME 1270
Db 1225 FSAFITYLOMKEMKVSIDIPQYSQVLVANCETLOEFVIALPQTRHSIALGSATEDKDSME 1284
Qy 1271 TDDCSRSRHRORQDVCVLGHLAKEICEVEDDEGDSMLQVTRRLPILPTLLTLEVSIRM 1330
Db 1285 TDDCSRSRHRORQDVCVLGHLAKEICEVEDDEGDSMLQVTRRLPILPTLLTLEVSIRM 1344
Qy 1331 KONLHFEATLHLITTLARTOOGATAVAGAGITQISICPLLSVYQSLSTNGTAQTPSARK 1390
Db 1345 KONLHFEATLHLITTLARTOOGATAVAGAGITQISICPLLSVYQSLSTNGTAQTPSARK 1404
Qy 1391 SLDAPSWGYYRLSWSLMEQLLKTLRYNPLPEALDPVGHQERTLOCLINAVTVQSLACL 1450
Db 1405 SLDAPSWGYYRLSWSLMEQLLKTLRYNPLPEALDPVGHQERTLOCLINAVTVQSLACL 1464
Qy 1451 EEAHTYVFIQLSNFMKEHMFHLBOLMRDIOVNLGYICOACTSILSHRRKMLQHLQKN 1510
Db 1465 EEAHTYVFIQLSNFMKEHMFHLBOLMRDIOVNLGYICOACTSILSHRRKMLQHLQKN 1524
Qy 1511 GDGCPSAVAGVORPPSAASAAPSSSKOPADTEASEOQALHTVOYGLKILISKTALR 1570
Db 1525 GDGCPSAVAGVORPPSAASAAPSSSKOPADTEASEOQALHTVOYGLKILISKTALR 1584
Qy 1571 HFTPDVCOILLDQSIDLAENYFLFALSTPTPTDSEVAPSPFGTLLATYVVALMMLGELDK 1630
Db 1585 HFTPDVCOILLDQSIDLAENYFLFALSTPTPTDSEVAPSPFGTLLATYVVALMMLGELDK 1644

```

```

Qy 1631 KKEPLTOAVGLSTQAEGRTRTLKSLMFTMENCFFYLLISOAMRYLRDPAVHRDQRMKE 1690
Db 1645 KKEPLTOAVGLSTQAEGRTRTLKSLMFTMENCFFYLLISOAMRYLRDPAVHRDQRMKE 1704
Qy 1691 LSSELSTLSSLSYFRRGAPSSPATGYLPSPOGKTSLSAASPSQEPILQLOVAFVRH 1750
Db 1705 LSSELSTLSSLSYFRRGAPSSPATGYLPSPOGKTSLSAASPSQEPILQLOVAFVRH 1764
Qy 1751 MOR 1753
Db 1765 MOR 1767

RESULT 22
US-10-719-385-23
; Sequence 23, Application US/10719385
; Publication No. US20040209284A1
; GENERAL INFORMATION:
; APPLICANT: O'Toole et al.
; TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis
; FILE REFERENCE: 22058-582
; CURRENT FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: PCT/US03/37339
; PRIOR FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: 60/428,094
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 23
; LENGTH: 1111
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-385-23

Query Match 58.5%, Score 5270, DB 17, Length 1111,
Best Local Similarity 92.8%, Pred. No. 0,
Matches 1031, Conservative 36, Mismatches 34, Indels 10, Gaps 1,

Qy 653 LSQMSAEGNAGGYGNILMSSEOPQGEYGVITAFRLITTLVKQGLSTQSGQLVPCW 712
Db 1 MTQWISAEGNAGGYGNILMSSEOPQGEYGVITAFRLITTLVKQGLSTQSGQLVPCW 60
Qy 713 FVLKEMLPSTHKRRYNSHGVRQIGCLIELHAHNLCHETDLSHSTPSLOPLCISL 772
Db 61 FVLKEMLPSTHKRRYNSHGVRQIGCLIELHAHNLCHETDLSHSTPSLOPLCISL 120
Qy 773 AYTEAGQTVNINIGVDTIDMMAAOPRSDGAEQGGQGLIKTVKLAFTSVNNVIRLK 832
Db 121 AYTEAGQTVNINIGVDTIDMMAAOPRSDGAEQGGQGLIKTVKLAFTSVNNVIRLK 180
Qy 833 PPSNVSPLEQALSHGAGNNLIATLAKTIYKHDPAFLRAIOLIKRLATVAPMSYTA 892
Db 181 PPSNVSPLEQALSHGAGNNLIATLAKTIYKHDPAFLRAIOLIKRLATVAPMSYTA 240
Qy 893 CLGNDAAIRDAFLTRLOSKEIDMRKIMLEPLTAVETOPGLIELFLNLVYKDSGSGS 952
Db 241 CLGNDAAIRDAFLTRLOSKEIDMRKIMLEPLTAVETOPGLIELFLNLVYKDSGSGS 300
Qy 953 KEFSLGMSCLHAYLELIDSQODRYWCPPLHRAAIAFLHALMODRDSAMLVLRTPK 1012
Db 301 KEFSLGMSCLHAYLELIDSQODRYWCPPLHRAAIAFLHALMODRDSAMLVLRTPK 360
Qy 1013 FWNELTSPFLGTSPSETSPSILETCALIMKICIEIYVVVKSIDQSLKDTLKKFS 1072
Db 361 FWNELTSPFLGTSPSETSPSILETCALIMKICIEIYVVVKSIDQSLKDTLKKFS 420
Qy 1073 EKRPAVSGYKSLAVHAETEGSSCTSLLEYOMLVASAMRMLIATTHADIMHLDVSV 1132
Db 421 EKRPAVSGYKSLAVHAETEGSSCTSLLEYOMLVASAMRMLIATTHADIMHLDVSV 480
Qy 1133 RROFLDVLDTGKALLVPAVNCILRGSMECTLLILLRQMKELGSVDEILGPLETEL 1192

```

```

Db      481 RROFLDVLDTGKALLVAASVNCRLGSMNCTLLILRLQMKELGAVKILGPLTEIL 540
Qy      1193 EGVQAOQOQMEKXKAKVFSAFITVLQMKEMKVSIDIPOYSOLVNVCTLOEEVIALFDQ 1252
Db      541 EGVQAOQOQMEKXKAKVFSAFITVLQMKELRGVIGIPYSOLVNVCTLOEEVIALFDQ 600
Qy      1253 TRHSLALGSATEDSDMETDDCSRRHRDQRDGVCLGLHLAKELCEVEDSDMLQVTR 1312
Db      601 TRHSLASSAEDSDMETDDCSRRHRDQRDGVCLGLHLAKELCEVEDSDMLQVTR 660
Qy      1313 RLPLPLTLTLTLVSLRMKQNLHFTTEATLTLTLTLTLTLTLTLTLTLTLTLTLTLTL 1372
Db      661 RLPLPLTLTLTLVSLRMKQNLHFTTEATLTLTLTLTLTLTLTLTLTLTLTLTLTLTL 720
Qy      1373 VYQSTNGTACTPBAASRSKLDAASMPGYRLSMSLMEQLTKLRYNPLPEALDVGVAH 1432
Db      721 VYQSTNGTACTPBAASRSKLDAASMPGYRLSMSLMEQLTKLRYNPLPEALDVGVAH 780
Qy      1433 RTLOCLINAVRTVQSLACLSEADHTVGFILQSLNFKEMHFFHLPOLMRDIQVNLGYLCOAC 1492
Db      781 RTLOCLINAVRTVQSLACLSEADHTVGFILQSLNFKEMHFFHLPOLMRDVQVNLGYLCOAC 840
Qy      1493 TSLHSRMTLOHYLQNNKNGDGLPSAVQAVQRPESA-----ASAPSSSKOPAD 1542
Db      841 TSLHSRMTLOHYLQNNKNGDGLPSAVQAVQRPESA-----ASAPSSSKOPAD 900
Qy      1543 TEASEQOALHTVOYGLKILSKTLAALHFTPDVQOILLDOSLDLAENPLFALSFTPT 1602
Db      901 TEASEQOALHTVOYGLKILSKTLAALHFTPDVQOILLDOSLDLAENPLFALSFTPT 960
Qy      1603 FDSEVAPFSGTLTATVNVVALNMLGELDKKKEPLTQAVGLSTQABSTRILKSLNFTWENC 1662
Db      961 FDSEVAPFSGTLTATVNVVALNMLGELDKKKEPLTQAVGLSTQABSTRILKSLNFTWENC 1020
Qy      1663 FYLLISQAMRYLRDPAVHRDKQRMKQELSEBLSLSSRYRRGAPSSPATGVLPSP 1722
Db      1021 FYLLISQAMRYLRDPAVHRDKQRMKQELSEBLSLSSRYRRGAPSSPATGVLPSP 1080
Qy      1723 QGKSTSLSKASPESEPLIOLVQAVFVRHMOR 1753
Db      1081 QGKSTSLSKASPESEPLIOLVQAVFVRHMOR 1111

```

## RESULT 23

```

US-10-719-385-24
; Sequence 24, Application US/10719385
; Publication No. US20040209284A1
; GENERAL INFORMATION:
; APPLICANT: O'Toole et al.
; TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis
; FILE REFERENCE: 22058-582
; CURRENT FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: PCT/US03/37339
; PRIOR FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: 60/428,094
; PRIOR FILING DATE: 2002-11-21
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 24
; LENGTH: 853
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-719-385-24

```

```

Query Match      45.4%; Score 4093; DB 17; Length 853;
Best Local Similarity 96.2%; Pred. No. 0;
Matches 810; Conservative 5; Mismatches 21; Indels 6; Gaps 1;
Qy      683 VTIAFLRLITLTVKQGLSTQSGIIVPCVMFVLKMLPSYHKWRNSGVRBOIGCLILE 742
Db      1 VTIAFLRLITLTVKQGLSTQSGIIVPCVMFVLKMLPSYHKWRNSGVRBOIGCLILE 60

```

```

Qy      743 LTHAILNLCHETDLHSHSTPSLOFLCICSLAYTEAGQVINIMIGVDTIDMVAAPRS 802
Db      61 LTHAILNLCHETDLHSHSTPSLOFLCICSLAYTEAGQVINIMIGVDTIDMVAAPRS 120
Qy      803 DGAEGGQGGQLLKTVKLASVTNNVRLKRPNSVSPLEQALSGHAGNNLIAYLAKY 862
Db      121 DGAEGGQGGQLLKTVKLASVTNNVRLKRPNSVSPLEQALSGHAGNNLIAYLAKY 180
Qy      863 IYHKHDPALRLAIQLKRLATAPMSVYACIGNDAAIADAFLTRLOSKIEDMRIKYMI 922
Db      181 IYHKHDPALRLAIQLKRLATAPMSVYACIGNDAAIADAFLTRLOSKIEDMRIKYMI 240
Qy      923 LEPFLTAVETOPGLIELFLNLLEVYDGSQSKFSLGMSCLHAYLELIDSGQODRYMCP 982
Db      241 LEPFLTAVETOPGLIELFLNLLEVYDGSQSKFSLGMSCLHAYLELIDSGQODRYMCP 300
Qy      983 LTHRAAIAFLHALMODRRDSAMVLRKPKFWENLISPLFGTSLSPSETSEPSLLETCAL 1042
Db      301 LTHRAAIAFLHALMODRRDSAMVLRKPKFWENLISPLFGTSLSPSETSEPSLLETCAL 360
Qy      1043 IMKILCEIYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVY 1102
Db      361 IMKILCEIYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVY 420
Qy      1103 EYQMLVSAMRMLLIATTHADIMHLTDSVVRQLFDVLDTGKALLVPSVNCRLGSM 1162
Db      421 EYQMLVSAMRMLLIATTHADIMHLTDSVVRQLFDVLDTGKALLVPSVNCRLGSM 480
Qy      1163 KCTLLILLLQMKRELGSVBEILGFLTEILEGVQAOQOQMEKXKAKVFSAFITVLQMK 1222
Db      481 KCTLLILLLQMKRELGSVBEILGFLTEILEGVQAOQOQMEKXKAKVFSAFITVLQMK 540
Qy      1223 MKVSDIPOYSOLVNVCTLOEEVIALFDQTRHSLALGSATEDSDMETDDCSRRHRDQ 1282
Db      541 MKVSDIPOYSOLVNVCTLOEEVIALFDQTRHSLALGSATEDSDMETDDCSRRHRDQ 600
Qy      1283 RDGVCVLGLHLAKELCEVEDSDMLQVTRRLPLTLTLTLTLTLTLTLTLTLTLTLTL 1342
Db      601 RDGVCVLGLHLAKELCEVEDSDMLQVTRRLPLTLTLTLTLTLTLTLTLTLTLTLTL 660
Qy      1343 LLLTLARTOOGATRVVAGAGITQSLCLPLLSYVQSLSTNGTACTPBAASRSKLDAASMPGYR 1402
Db      661 LLLTLARTOOGATRVVAGAGITQSLCLPLLSYVQSLSTNGTACTPBAASRSKLDAASMPGYR 720
Qy      1403 LSNFMEQLKTLRYNPLPEALDPVGVHOBRTLOCLINAVTVOSLACLSEADHTVGFILQ 1462
Db      721 LSNFMEQLKTLRYNPLPEALDPVGVHOBRTLOCLINAVTVOSLACLSEADHTVGFILQ 780
Qy      1463 LSNFMEQHFFHLPOLMRDIQVNLGYLCOACTSLHSRKMLOHYLQNNKNGDGLPSAVQAV 1522
Db      781 LSNFMEQHFFHLPOLMRDIQV-----ADDGVLESVMLGDRBAVSHMGTPELQDVP 834
Qy      1523 QR 1524
Db      835 ER 836

```

## RESULT 24

```

US-10-719-385-25
; Sequence 25, Application US/10719385
; Publication No. US20040209284A1
; GENERAL INFORMATION:
; APPLICANT: O'Toole et al.
; TITLE OF INVENTION: Composition and Method for Treating Lupus Nephritis
; FILE REFERENCE: 22058-582
; CURRENT FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: PCT/US03/37339
; PRIOR FILING DATE: 2003-11-21
; PRIOR APPLICATION NUMBER: 60/428,094
; PRIOR FILING DATE: 2002-11-21
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: Patentin Ver. 2.1

```

SEQ ID NO 25  
LENGTH: 525  
TYPE: PRT  
ORGANISM: Mus musculus  
US-10-719-385-25

Query Match 26.9%; Score 2420; DB 17; Length 525;  
Best Local Similarity 91.6%; Pred. No. 2.3e-210;  
Matches 478; Conservative 13; Mismatches 21; Indels 10; Gaps 1;

QY 1242 LQEVIALFDQTRSLALGSAEDKDSMETDSCSRSHRDQDVCVLGLHAKELCEVD 1301  
DB 4 LQEVIALFDQTRSLALGSAEDKDSMETDSCSRSHRDQDVCVLGLHAKELCEVD 63  
QY 1302 EDGDSMLQVTRRLPILPTLLTLEVSLEMKONLHFTTEATLLTLTARTQCATAVAGAG 1361  
DB 64 EDGDSMLQVTRRLPILPTLLTLEVSLEMKONLHFTTEATLLTLTARTQCATAVAGAG 123  
QY 1362 ITOSICPLLSVGLSTNGTAQTPSASRKSLLDAPSWPVRLSMLEOLKTLRYNPLP 1421  
DB 124 ITOSICPLLSVGLSTNGTAQTPSASRKSLLDAPSWPVRLSMLEOLKTLRYNPLP 183  
QY 1422 EALDFVGHQERTIQCLNAVTVQSLACLEADHTVGFILQLSNPKEMHFLPOLMRDI 1481  
DB 184 EALDFVGHQERTIQCLNAVTVQSLACLEADHTVGFILQLSNPKEMHFLPOLMRDI 243  
QY 1482 QVNIGYLCQACTSLHSHKMLQHTLQNKNGDGLPSAQAQVQRPSPA-----ASA 1531  
DB 244 QVNIGYLCQACTSLHSHKMLQHTLQNKNGDGLPSAQAQVQRPSPA-----ASA 303  
QY 1532 APSSKQPAADTESEQALHTVQGLIKITLTAALRHFTPDVCOILLDLSLEVN 1591  
DB 304 APSSKQPAADTESEQALHTVQGLIKITLTAALRHFTPDVCOILLDLSLEVN 363  
QY 1592 FLFALSTPTPTDSEVAPFSGTLLATVVALNMLGELDKKEPLTOAVGSLTOAGRTTL 1651  
DB 364 FLFALSTPTPTDSEVAPFSGTLLATVVALNMLGELDKKEPLTOAVGSLTOAGRTTL 423  
QY 1652 KSLIMFTWENCFYLLISQAMRYLDDPAVHPDKQRMKQELSELSTLSSIRYFRGAP 1711  
DB 424 KSLIMFTWENCFYLLISQAMRYLDDPAVHPDKQRMKQELSELSTLSSIRYFRGAP 483  
QY 1712 SSPATGVLPSPQGSTLSKASPSQEPILQVQAFVHMOR 1753  
DB 484 SSPATGVLPSPQGSTLSKASPSQEPILQVQAFVHMOR 525

## RESULT 25

US-10-437-963-135530  
Sequence 135530, Application US/10437963  
Publication No. US20040123343A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa, Thomas J.  
APPLICANT: Kovalic, David K.  
APPLICANT: Zhou, Yihua  
APPLICANT: Cao, Yongwei  
APPLICANT: Wu, Wei  
APPLICANT: Boukharov, Andrey A.  
APPLICANT: Barbazuk, Brad  
APPLICANT: Li, Ping  
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21(53221)B  
CURRENT FILING DATE: 2003-05-14  
NUMBER OF SEQ ID NOS: 204966  
SEQ ID NO 135530  
LENGTH: 2026  
TYPE: PRT  
ORGANISM: Oryza sativa  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (1)..(2026)

OTHER INFORMATION: unsure at all Xaa locations  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT4530\_37198C.1.pep  
US-10-437-963-135530

Query Match 4.1%; Score 369.5; DB 16; Length 2026;  
Best Local Similarity 18.1%; Pred. No. 9.9e-23;  
Matches 383; Conservative 327; Mismatches 653; Indels 755; Gaps 98;

QY 25 LGRSALRELSQIEA---ELNKRRLLEGSLYRPPSPSSAEKANKANDVAS-----73  
DB 45 LDRRAAASPVPDIALAIRAHAMLRGVSMPFKPSDASGALDASBVGHEHRLAVKP 104  
QY 74 PLKELGRISKFGGLDESGVQLIQYQEDYRGTRSVKTVLQDSEOSQALLIKIDY 133  
DB 105 ELKAAALRLSKOMLDEVOSTYIVL-----KTSFTPTALVAD---TEEFLLVSVQY 154  
QY 134 YEERTCILRCVLIHLTFYFODERHRYVEYADCVDKELVSKYQOFEE-----LYKTE 188  
DB 155 YLERQCLIKCIRRIFFVHND-----CSDSIDAVEBAGVLVREEVQRLSIVRSL 206  
QY 189 APYETHGNLMTERQVSRWVQCLREOSMLLEIFLYIAYEPASD---LLVLTMYKE 245  
DB 207 ASAFSVKGG---AEITLIS-WLEETLIRINLIFDIIFLEF---YDNLSCRNGGLMIMLCIFKD 262  
QY 246 QGFSRQTNRLVD-ETMDPVRDRIGYPSA---LIVGMDIESLHKALDQD---RRELIH 298  
DB 263 MLSSGYDVKFAVSEAKNSF---HYAKAQLLITLQTDPEBSLRWVRDVEPPSGYS 318  
QY 299 QFADGLICQDMCLMTFQDIPRHAPVLLAMALLRHTLNEETSVMKRGTAIQLNV 358  
DB 319 TFSVVDLLEMDVEBSKLPFAVAVESGPLILAMAF-----LCVMSLPGSNVNLGP 369  
QY 359 FOYVTRLLQSLAGSANDCTTSYACMCYGLSFVLSLHHTLNGQODIDTACEVLADP 418  
DB 370 VSGRGLIRFTIS-----AFVASY-----388  
QY 419 SLPELFWCTEPTSGIGIILDSVCGMFFHLSPILLQ---LRALVSGKSTAKKVSFLDKS 476  
DB 389 ---EISYQTEBS-LGMILNLTCEVYDEESLQCMQFMDKDSFIDGPR---FNILEMKN 440  
QY 477 FYNELY---KRPKPHVISHEDGTLWRQTPKLLYPLGQGTNLRIPQGTGVQVM---LDDRAY 532  
DB 441 GVTLLYAVPRSDTDNMYNDQI---EHSPIISIGIEGT---IGGSGIYLKXLEDVA 495  
QY 533 LVRWEYSYSWTLFTCEIEMLLHAVSTADVIQHCQRYKPIIDLVHXYISTDLSTADCLP 592  
DB 496 LVRWE-----DLCLALLHADKSLAVQASQN---LGYIDKVRIDIAKIFC---537  
QY 593 ITSRIYMLQRLT---VISPVVDYIASCNCLTYLAARNAKPMKMTDLRHGTFLPVHAPV 650  
DB 538 ---TSIFKYVEDPNNACVMSKTLGMLAEMLSCV-----PYNHVNVALDCGP---FTTG- 584  
QY 651 SSLSOMTSAEGMNAAGVGNLLMNSQPOGEGVTIAPRLTLTVKQL-----GSTSQ 705  
DB 585 ---SGVASDMLSLGALARKLPATSBDSGDS-----SLTTYVDFAIQLVLRGAADD 635  
QY 706 GLVPCVMFLKMLPSYHKRYSNH-----GVREIGCLILEP---743  
DB 636 IISFIIPSVQYIWMNHNWKKYSRWKITLKYFDLVKSCIQVKSFSKGLGIIMETIL 695  
QY 744 ---IHALL-NLCHETDL---HSHRPSQFLCISGLATBAGQYIYNIMGIVPDI 793  
DB 696 YDSSHSVYLHILMSYQLLHSHSGSYCHDKDI-----EDQLVLCGCFDVF 744  
QY 794 MVMAAQPRSDGAEQOGOLIKTV---KLAFSVTNVVRILRPPSNVVSPLRQALSOHGAH 851  
DB 745 YMLSNLP-----EREKSKMLRKYGSKSARYYGFFVLVKTIA-----FRH 786  
QY 852 GNNILAVLAKY-----YHKHDPALPRLAIQLIKRLATVAPMSVTAACLAND---897  
DB 787 SCNIIDVLSFSPSIEWRTCIIFH-----LQLAFIKILS-----QSHVACSNSEDNN 834



```

QY 898 -----AAA-----IRDAFLTRLOS- 911
Db 835 RTSNKPMOTDNTDIFLKAIVAAARVFSMLCTAVAKQPOLMENNAFVFNVSSEIWRLOTS 894
QY 912 -----KEDMRKXMLEFLTVAVETOP-----GLIFLFLNLEVKDGS----- 949
Db 895 ISCIIDEVDKNEVVAIFNLSSARQADVALVWLRYHLPEOOSTRAQADSNSAH 954
QY 950 -DGSKEFSLGMSCLHVALELI-----DSQOQDRYWCPEPLHRAAIAFLHLMQDRDS 1002
Db 955 EQSKYFVLNPSGSPRLVEQILGIGSTELMDR---SPSLSSVLDLALHBSGAQF 1011
QY 1003 AMLV--LTKRKPENLTPSPFGL--SPSET---SEPSILETCAL--INKIICLEY 1053
Db 1012 IYILEKRSRTFMENLSCCIRAAPASYPIDSVETDEKKSIRYCCIGCTIPEINSEYFL 1071
QY 1054 VVK-----GSLDQ-----SLDYLTKKSEIEKFAVWS 1080
Db 1072 QGKLLTETKTSDDPAVGSKEQKPSVADCPSDIYLKMPDSTTMEDLVNHLSS-----N 1124
QY 1081 GYVKSIAVHVAETEGSSCT-----SLEYQMLVSA-----WRMLL 1115
Db 1125 GYQNDL-IHRAKAASCRITILITLTSAGDTGSLSFSLVKKIQLISSKULTMCGARTLL 1183
QY 1116 I-----IAT-----HADIMHLTDSVVRQLFLDVLDTYKALLVPASVNCRLGS 1161
Db 1184 CGGPRIASGTGQVWTHRRHGDRAQSTDA-----NAFGLQGVVTF--CI----- 1227
QY 1162 MKCTLLILLRQWKRELGSVE-----ILGPLTEIL----- 1192
Db 1228 --CSSTVITCPSWLOMLIFMDEPNFLHDIMLKLIHAELEGRTTPPEFQELLCFLELF 1285
QY 1193 -----EGVLQAD-----OQLMEK----- 1205
Db 1286 KVFENHSEQLOKSPRANGVSLPVDVPHIRDELGLHLMNSDMKYKAVAKMDIMEKA 1345
QY 1206 -----TKAKVSAFTVLQW---KEMKVS DIP-----QYSLVNLVNCETLOEV 1246
Db 1346 NLMKQVDAKLALRSLFITFLSVCTGSSYKKGFLPGGIGSITTTQSAVRCACKSLQSAV 1405
QY 1247 IALPDQTHS-----LALGATBKDS-----MENDDQSR 1277
Db 1406 DSLPEVDNSGVLPPLSGVLELITTRILLDHAKQKSSRHHYPVIVLIMKTSIGASTS 1465
QY 1278 -----RHRDQDG-----VCYLGLHAKEL 1297
Db 1466 FLFLMPPSPALKQPVKSLVLLSLPFIKYKQDMKQSDVNI FGLSLSLSPVL 1525
QY 1298 CEVDEDDG-----SWLQVTRRLPIELTTLTLEVLKMKQW--LHFTERTL 1341
Db 1526 CKLAESREYFDLAIASMDIILKGPLSNVWVPILOKHP-RLQVILLOKQSGALICTQVTL 1584
QY 1342 HLLTLARTOGATAVAGAGITQSLCLPLSVYOLSTNGTQOTPSASRKSIDAPSWGCV 1401
Db 1585 NFLITMERTKGAALIKLOSANIPAFIKV-LIS--QMSLDDSCLRSLSLQYTDVKIW---- 1637
QY 1402 RLSSMLMEQLKTL-----RNFPLPEALDF-----YGVHOERTLOCIINA 1440
Db 1538 GLGLAIVSLNHCMDDDISRNSVANSTISFLSGVPLMSSYLSAQSVAVTHQSKRTILOK 1697
QY 1441 VRTYQSLACLEADHYTGFIQLQSNFMKEWFFHLPOLMRDIQVNLGYLQACITSLHSRK 1500
Db 1698 SQT--SISALSLTER--ILTLCLIAKXHPRPDTGKKEVDSEI-----RE 1738
QY 1501 MLQHYL-----QMKNGD-----GLPSAVAQRV---QRRP-----SAAS 1530
Db 1739 ITHLLFLFISRGSERTGSPWNWLSFGCPPIIKESMKINEEPPLIRSKYGFRRPAACTL 1798
QY 1531 AAPSSSKQF-----AADTEASEQO-----ALHTVOYGLKTL--SKTLAAL 1569
Db 1799 STPEVSGPNNAGSLVIRDKNPAQSDSMKQRTFEMTLAVQVYIRAFILMKRLGQAKNAV 1858
QY 1570 R-----HF-----TPVQCILLDQSLDLAENFLFALSTTPTTPESEVAPSPGT- 1613

```

```

Db 1859 FRARELEFLDLAPPELMPEDILHGLQD-----VSVIVTEVLEAVSTALNTE 1907
QY 1614 -----LATVNVALNM 1624
Db 1908 TERVQCLLVLETSLYM 1925

```

## RESULT 26

```

US-10-106-698-5917
Sequence 5917, Application US/10106698
Publication No. US20030109690A1
GENERAL INFORMATION:

```

```

APPLICANT: Ruben et al.
TITLE OF INVENTION: Colon and Colon Cancer Associated Polynucleotides and Polypeptides
FILE REFERENCE: PA005P1

```

```

CURRENT APPLICATION NUMBER: US/10/106,698

```

```

PRIOR FILING DATE: 2002-03-27

```

```

PRIOR APPLICATION NUMBER: PCT/US00/26524

```

```

PRIOR FILING DATE: 2000-09-28

```

```

PRIOR FILING DATE: 1999-09-29

```

```

PRIOR APPLICATION NUMBER: US 60/163,280

```

```

PRIOR FILING DATE: 1999-11-03

```

```

NUMBER OF SEQ ID NOS: 8564

```

```

SOFTWARE: PatentIn Ver. 3.0

```

```

SEQ ID NO 5917

```

```

LENGTH: 63

```

```

TYPE: PRT

```

```

ORGANISM: Homo sapiens

```

```

FEATURE:

```

```

NAME/KEY: MISC FEATURE

```

```

LOCATION: (36)

```

```

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

```

```

US-10-106-698-5917

```

```

Query Match 3.4%; Score 305; DB 14; Length 63;

```

```

Best local similarity 96.6%; Pred. No. 1,7e-19;

```

```

Matches 56; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 953 KEPSLGWMSCLHVALELIDSOQODRWYCPPLHRAAIAFLHLMQDRDSAMVLRK 1010

```

```

Db 6 QERSLGMSCLHVALELIDSOQODRWYCPPLHRAAIAFLHLMQDRDSAMVLRK 63

```

## RESULT 27

```

US-10-437-963-173637
Sequence 173637, Application US/10437963
Publication No. US2004012343A1
GENERAL INFORMATION:

```

```

APPLICANT: La Rosa, Thomas J.

```

```

APPLICANT: Kovalic, David K.

```

```

APPLICANT: Zhou, Yihua

```

```

APPLICANT: Cao, Yongwei

```

```

APPLICANT: Wu, Wei

```

```

APPLICANT: Boukharov, Andrey A.

```

```

APPLICANT: Barbazuk, Brad

```

```

APPLICANT: Li, Ping

```

```

TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With

```

```

FILE REFERENCE: 38-21(53221)B

```

```

CURRENT APPLICATION NUMBER: US/10/437,963

```

```

CURRENT FILING DATE: 2003-05-14

```

```

NUMBER OF SEQ ID NOS: 204966

```

```

SEQ ID NO 173637

```

```

LENGTH: 1745

```

```

TYPE: PRT

```

```

ORGANISM: Oryza sativa

```

```

FEATURE:

```

```

OTHER INFORMATION: Clone ID: PAT_MRT4530_71655C.1.pep

```

```

US-10-437-963-173637
Query Match 2.7%; Score 245; DB 16; Length 1745;

```



Best Local Similarity 18.8%; Pred. No. 1.6e-11;  
Matches 338; Conservative 279; Mismatches 623; Indels 554; Gaps 88;

Qy 270 GYFALLIVEGMDI-----ESLHKALDDRREHQAQDGLICQDMCLMTFGDIPH 322  
Db 18 GVFFSLVYTFVSDALSTSHKPSLSSSDSSFRDHEU-----VMSDNMLTEGFGV-- 71  
Qy 323 HAPVLLAALLRHILNPEETSSVAKIGTALQANVPYLRLLQSLASGNDCTSTAC 382  
Db 72 ---VRLAAV--HMLLTQDRSSA-----RDTLTSSSRDVTDIWAC 106  
Qy 383 M---CVYGLSFLVLSLEHTLGNQDIIIDACEVLADPSLPFLFWG-----TEP 429  
Db 107 LEICRQMSFOFL-----QERIMQTRAYKNDDEDIVMTGTMLMLMCFLSHP 155  
Qy 430 TSGGLIIDSVCNCFPHLLSLQLLRALVSGSKAKV-----YSLDKMSFYNELYKH 484  
Db 156 TSDKI--KEIKEKTMALSPYGSIRDRHEDPSRTEQIQPTNPFISSLLELVREIYOK 213  
Qy 485 KPHVISHEDGTLW-----RQTPKLYPLGGQNNLRIPQGTVCQV--MLDDRAX- 532  
Db 214 EPELVHGER--LWTFVISAGEDHTNLTLLVAFGLSLSTLASSEVGAAYVELLOQKYR 271  
Qy 533 LVRWEYSVSWTLFTC---ELEMILHYSTADVIQH-----CQRKPIIDLVHKVISTD 583  
Db 272 SLGMS-----TLFDCLSIYEKFKESLOSASHVPEEPADQALVSYLAVLOKVENG 325  
Qy 584 LSIADCLLPITSR-----IYMLQRLTVVISPVDVIASCVNCLTVLAARNA---KVM 634  
Db 326 NT-----TERRKWPDIIEPLFKLSYENVPPF-LGALRNSITARTKVSPLKDAIM 376  
Qy 635 TDLNHTGFLPRVAHVVSLSQMSAEGNAGYGNL---LANSQPOEYGTIAFLRLI 691  
Db 377 SYLQOYD-LPVATPELQO-----HNATOIYMRLENEVEEMRESYSTISFLMW 426  
Qy 692 TTLVKGOLG-STOSQGLVPCWFLKEMLPYHKRKYNSHGVRQEOIGCLLELHAIINL 750  
Db 427 NALLAEENINIDKGRFMEGIRKFEYEDVFPFPQOAYADPREKELAAVACLEHFMVLSM 486  
Qy 751 ---CHETDLH-----SSHT-----PSIQFL--CICSLAYT----- 775  
Db 487 YDIKODDIYAAMNAGSPSTSHASIDROLVLELKKSCTLSLPHVLSAIPVIKHNDP 546  
Qy 776 ----EAGCTVINIMIGVDTIDMWAAAPRSDGAGGQGLITITVGLASVTNNVR- 830  
Db 547 MSGKVAFNINIVISVGVDTL-----INERTTQYGLLEKTVHLSFEIFLWVER 597  
Qy 831 -----LKPPSNVSPLEQALSHOAGHNNLIAYLAKYIYKHDPALPRALIAOLKRLA 883  
Db 598 DLVLADVRP--LYQPLDVLVAQNH--IALLERFYDYLPJOQCSIKIMGLIS 650  
Qy 884 TVAPMSVYACLGNDAA-AIRDAFLTRLQSKIEDNR-----IKVMLIEPLTVAV-ETQ 933  
Db 651 SRIYGLVQLLKADVAKSVIEDYAACLBFRPDPQAIENITKODVGLLQILLIDNICPA 710  
Qy 934 PGLIEFLANLVKQSDGSKPSL---GMSCLHAYTELDSQOQDXYWCPPLHRAI 989  
Db 711 PNITHTLIRFDV---NGSIRTVAKPKSHVSCLEKTIIDNEKVKTPD--INALHERGF 764  
Qy 990 AFLHALMQDRDSAMLVLRKPKFMENTSPFLGTLSPSESESIETCALMIKICL 1049  
Db 765 QLTLELCID-----PL-----TCGPVNDLST 786  
Qy 1050 EIVY-----VVKSGLDOSLQDTLKKFSIERKFAVWSGVKSLAVHVAETGS 1096  
Db 787 TKYQFSGHVGITIGSPLPRNNOSLRISW---LHER--AMLLKMLALHVSDISS 840  
Qy 1097 ----SCTSLLEYOMVAMRMILLIATTHADIMHLLTDSVVRQLFD-----VLDGKA 1146  
Db 841 LYRSC-----LAILCHTFG--HCAENLRSANLLQSCSSNLANNGKV 882  
Qy 1147 LLLV-----PASVNCRLSGSMKCTLLILLRQKRELGSV-----DEIL 1185

Db 883 LDLEVVQFRCRPTSIKYPQLSNLRLES---KIEILRNSATSEFGVYYSERGRLI 939  
Qy 1186 GPVLEIEGVQAOQ---QMEKTKAKVPSAFITV-----LOMKEMKXSDIQY 1231  
Db 940 -DLAFAHEKLLQMSQLNPNLOSEKSELKESFQOMKMAWYNNKZLEBOAQLHMLTGW 998  
Qy 1232 SOLVIANVCEITQERIALFDQTRSL-----ALGSATEDKDSMETDSCS----- 1275  
Db 999 SOIV---EIVASRMSILBERSHLFELDLASISAT-----TSPDCSVMAIYLITVA 1048  
Qy 1276 ---RSRHRDOR-----DGVCVLGLHLAKELCEV-----DEGD-- 1305  
Db 1049 LTCMAKRLDERFICPTGADSAVTCLDIISAKQSNACSTSLPFLTMAILRNESSETLR 1108  
Qy 1306 -----SWLOVTRL-----PIELTLITLB----- 1325  
Db 1109 RROYALLSTFYQCRSTLSDSVPPVIRFLLEBQEGDDDELQKVKLEONEELARSNFS 1168  
Qy 1326 -----VSLRMKONLHPTA-----TLHLTLTARTQOGATVAVAGITQSICLP 1369  
Db 1169 IIRKBAQAVIDLVAKDAIHSEBAGKAISFYVLDLSLIDHDKYFLNOIQSRGILRS-CLS 1227  
Qy 1370 LLSVYQSTNGTAQTPSASR-KSLDAP-----SWPGYRLS----- 1404  
Db 1228 DVNNY-LSKKASFPSESSQRCFTIDAQSLLRISHYGHKGSQILLSMGALHNLSCNL 1286  
Qy 1405 MSLMBQLIKTRYNFLPEALDFGVGHQERTLOCLNAVTVOSLACLEBADHTVGTIOLS 1464  
Db 1287 MGSQKKANSRLNSVVKERAGEIDKRSLTAPIIRIYTSFTSL-VDSAD----- 1334  
Qy 1465 NPMKEMHFL-----POLMRDQVNLGYL-----CQAC-----TSLHSR 1499  
Db 1335 -FLEAWYELIAYGTSARTDCPMLGFDL---CLIVYDADSWIGCKGCPGHAMFTINSL 1391  
Qy 1500 KMLQHYLQNK---GDGLPSAVAOVRPSPASAAPSXKOPAD-----TEASEQQA 1550  
Db 1392 VKIDRPLEDENLARGDNVKKVMTAIFRLLEABAP---QPTDRRLKMWMSQATKNQ 1447  
Qy 1551 LHTVOYGLIKTL-SKTLAALHFTPD---VQOILLDSL-----DLAEYVFLRALSTT 1600  
Db 1448 -KEVRKGINSLVKKIYVELVDFAKHQFVNIILRESISANIFNERLMMVVISLQKV 1506  
Qy 1601 PTPDSEVAPSRGTLTATVNAVLNMLGELDKKEPLTQAVGSTAEGRTLSKSLMPTME 1660  
Db 1507 WAYENDECSFVQDLFSM---MQFLPSLIDISLAMPQS---SNMIENOKS--ELIVFL- 1557  
Qy 1661 NCF-----YLLISOAMRYLRDPAVHPROKMKOELSELSTLSLSRYFRGA 1710  
Db 1558 -CFSLISLTVYLARKKMRPQISYDPTTEGGOQPTLQILSDLNSITVAMERVA 1610

RESULT 28  
US-10-437-963-154118  
Sequence 154118, Application US/10437963  
Publication No. US20040123343A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa, Thomas J.  
APPLICANT: Kovalic, David K.  
APPLICANT: Zhou, Yihua  
APPLICANT: Cao, Yongwei  
APPLICANT: Wu, Wei  
APPLICANT: Boukharov, Andrey A.  
APPLICANT: Barbazuk, Brad  
APPLICANT: Li, Ping  
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
FILE REFERENCE: 38-21(53221)B  
CURRENT APPLICATION NUMBER: US/10/437, 963  
NUMBER OF SEQ ID NOS: 2003-05-14  
SEQ ID NO 154118  
LENGTH: 2905  
TYPE: PRT

ORGANISM: *Oryza sativa*FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT4530\_54008C.1.pep  
US-10-437-963-154118Query Match 2.0%; Score 176.5; DB 16; Length 2905;  
Best Local Similarity 18.8%; Pred. No. 6,8e-05;  
Matches 323; Conservative 249; Mismatches 599; Indels 543; Gaps 86;

12 SPCSSRLMTLLGRSALRELSQLIABLNKWRLLGLSY-----KPPSSAK 64  
40 SFCR-----LAKRT-ANLKPALPASATWPLMALTKCVADISGKSPKSA 90  
65 -----VKANKVASPLKGLRISKEFLGDEOSVQLQCLQ 102  
91 RMLRLVACADFRLSGSLILISVAKKLFSHIWEVVDABSL-----TEYSLIKQLITV 147  
103 EDYGRDSTVKTLODEROSQALIL--KIADYEE-----RTCLIRCVAILTFYQ 152  
148 KEYR-----YQMPRTYSSFFVLVYMKKVGATGPEKISNQASKEESFRCLTLHVLLE 200  
153 DERHY-----RVEYADCVDLKELVSKTRQPEELYKT-----EA 189  
201 NPPGDPIMREESLFIYAKVQLELARAPEILKLDVILIKELQMFNTGAGLWCEA 260  
190 PWEHNGMLMTERQVSRWFVOCLEOSMLLEIIFYAYFEMAPSDLLVTFMKEQFG 249  
261 PRDEKAGSL-----RCFOELMDLSTALYLA-Y-KTPRTSHNKKLKT----- 304  
250 SKQTRHLVDETMDFPVDNRIGYFSAIIVEGMIDISLHKCALDREHLOFADQLICOD 309  
305 -----HMTAIMDS-----LLRGLSVMS-----GSLC-- 326  
310 MDCLMLTGGDIPHAHPVLLAALLAHTLNPEETSSVAKIGTALQLVFO----YLTRL 365  
327 --CLHKYQ-YKVDKSLINMF-----KATCGKLSVLSNANAAR-FQDSILMLIRV 374  
366 LOS-----LASGNDCTSTACMCVYGLSFVLTSLHTL--GNODIIDTCEVLADPSL 420  
375 LKEPALLMSAKRDP-----PFWLINDEMSLBEGYODWTNSLIHAL----- 417  
421 PELFWGTEPTSGLGILDSVCGMPHLL-----SPLLQRLALVSGSTAKK 467  
418 -PLF-----STTALVDSVLRLGEMIMRDVHASFVSEDTWDQIRKQPS--SSTLYF 469  
468 VYSELDKMSFY-----NELYKHK-----PHDIVSHEDGTLMRQT-----PKLLYPL-GEQT 513  
470 IACYFSKIGFOGDVNSIFIRKNLLRSTFELVHSGFSLNEQSVLMIPEITFSLCAGFS 529  
514 NLRIPOGTVGQV-----MLDDBAYLVWEXYSWTLFTCEIR--MLLHVSTADVI 563  
530 SPVINSATSQLFGECKNLKDKCMWSHEELGYSVETISEINLBEPYAKMEKSGRAHP 589  
564 QHCQRVKPI-----IDLVHKVISTDL--SIADLPLITSRIYML 601  
590 RHIG--QELLELELFTKGFMSNKKQFEKVDLPNLVYACSLVNCILHCSL--LSRVF-- 642  
602 QRLTVISPPVDVLASVNCCLVLAARPAKWTDLRTG-----FLPFWAHP 649  
643 EEKSSFLOVMDLVYNAIKHIVSVMKHAELSHGLTNVLGVPDITGSALSSPFSFMSP 702  
650 VSS-----LSQMISAEQNNAGGYGNLMSSEQPOGEYGTIAFLRLIT 692  
703 LFLSLMRVNNKTSVSLVTELDLELLELALISELPAQSSSLMNNPFDGNPBMKMLPVFVSLSE 762  
693 TL-----VKQQL-----GSTQSQGLV-----CVMFVLEKMLPS 721  
763 DLHPDNRSSFYVMDLMTGSEVPVTAAGSGSMGDISPRLECKLELVICISTEFSVAP- 821  
722 YHKMR--YN--SHGVRQIGCLILELHAILNLCHETDLHSHRPSLQ--FLICISLAY 774  
822 LHTWEVLNLADKEVNIRVQAIL-----DELCTNISASSSSLSASVQDVFLDLANKA- 874

775 TEAGQT-----VINMGIGVDTT--DMVWAAGPRSDGAGGCGQLLITVLAFL 822  
875 TEISFTMPERIRLINCISNTIYLPVDAQMT-----GHLL-----SML 914  
823 SVTNVNRILKPPSNVSPLEQALSQGHAGNMLIAVL-AKYI-YHKHDPALPRLAIQLLK 880  
915 HDTYRRLVYARIV-VLQIWECHSELPHDVSSSGVWVQGSNNPVAR----- 966  
881 RLATVAPMSVYACIGNDAAIRDAFLT--RLQSKIEMRIKMWILEFLTVAVETPGGLIE 938  
967 EVLAIGQPV-----PIIETALITIAHSLOSSEDIIEVECVMAVAIAIE--PSQRE 1016  
939 LFLNLVYKDSGDSKERSLGMWCSLHVLLEIDSQQDRY--WCPPLHRAAIAFLHAL 995  
1017 L-----TYALPDSVSKRLNAPSRSKIDOLIGFLFR----- 1048  
996 WQDRDSAMLVLRK-----PK-FWENTLSPFLGTLSPPSETPSILETCALIMK 1045  
1049 WVACEVSLVSEVKEHFGFSDPKNFIEHCSWLSFLIRGETADLWLS-----K 1102  
1046 IICLEIYVYVKGSLDQSLKDTLKKFSIEKRPAYSGVYKSLAVAVAEBSGSSCTSLLEY- 1104  
1103 ILSQPLSAVYKGYF-----VPIFGLCIARYSGGEKOLA-----ETVLCESLQLG 1149  
1105 -----QMLVAMRMLLIAT--HADIMHLDTSVVRQL-----FHDVLDT 1144  
1150 EISPERDDLKKGWISIVGFLSVSSSSROPELPHSKAVHNSVKTVDGFDMDNGN 1209  
1145 KALLVPASVNCRLSGMKCTLLILLRQWRELSGVEILGPI--TEILEGVLOAQOL 1202  
1210 LAETVVIDKINIFR--SDRVKFLAIHQIADASHPRHRIHRLCAIEVLIDVL----- 1261  
1203 MEKTKAVPFAFTVLOMKEMKVSIDIPOYQVLNVCTEOEVIALPD--QTRHSIAL 1259  
1262 --GHRVALYSTWFIYIIGSVYIQROPLOQ-----CNTLSKLLAFNSCSTETDAAL 1314  
1260 GSATEDKDSMETDCC-----SRGRHRODQGVCLGLHLAKECEVEDSDSWLQV 1310  
1315 GKOQDLEPLPSPCLQIDQVHNASLSGYDSR-----DQTLKF 1353  
1311 TRRLPIPT--LITLLEVSIRMKONLHFTBATLHLLTLARTOQATAVAGITQSIC 1367  
1354 VSRAPHLPPELFLSSSSVA-----NEASSVL-----ADFSRAGIS-- 1390  
1368 LPLISVYOL--SYNGAQTSPSARSKSLDAPSNPGVYRLMSIMEQLKTRRYNLEAL 1444  
1391 ---DVHOVIFHPVNLTEKYPLOPHISKE--DKLPSDYGISDDILVGLLKLACLSDESA 1445  
1425 PFGVHOERTLO--CINAVRTQSLACLEADHTV 1457  
1446 EIIDV-ISRTRLRGVLSYRAGLNLHCLDSIDRSL 1478

RESULT 29  
US-10-128-714-8246  
Sequence 8246, Application US/10128714  
Publication No. US20030119013A1  
GENERAL INFORMATION:  
APPLICANT: Jhang, Bo  
APPLICANT: Hu, Wenqi  
APPLICANT: Tishkoff, Daniel  
APPLICANT: Zamudio, Carlos  
APPLICANT: Erosbkin, Alexey M  
APPLICANT: Lemieux, Sebastien M  
TITLE OF INVENTION: Identification of Essential Genes in *Aspergillus fumigatus* and  
FILE REFERENCE: 10182-018-999  
CURRENT APPLICATION NUMBER: US/10/128, 714  
CURRENT FILING DATE: 2002-04-23  
PRIOR APPLICATION NUMBER: US 60/285, 697  
PRIOR FILING DATE: 2001-04-23  
PRIOR APPLICATION NUMBER: US 60/287, 066  
PRIOR FILING DATE: 2001-04-27

PRIOR APPLICATION NUMBER: US 60/295,890  
PRIOR FILING DATE: 2001-06-05  
PRIOR APPLICATION NUMBER: US 60/303,899  
PRIOR FILING DATE: 2001-07-09  
PRIOR APPLICATION NUMBER: US 60/316,362  
PRIOR FILING DATE: 2001-08-31  
NUMBER OF SEQ ID NOS: 8603  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 8246  
LENGTH: 1676  
TYPE: PRT  
ORGANISM: Aspergillus fumigatus  
FEATURE:  
NAME/KEY: MISC FEATURE  
LOCATION: (1565)  
OTHER INFORMATION: X= any amino acid  
US-10-128-714-8246

Query Match 1.9%; Score 175.5; DB 14; Length 1676;  
Best Local Similarity 17.9%; Pred. No. 3.2e-05;  
Matches 325; Conservative 282; Mismatches 665; Indels 543; Gaps 84;

QY 25 LGRSALBELSQIEALELNKH---WRRLLEGLSYKRP-PSPSSAEKVKANK-----DVA 72  
DB 20 LQDSSFVNIERLRYELETHIEDFKLKD-----KSKNNTSRQAVLSGKITVDVEYSIN 74  
QY 73 SPLKELGIRIKFGLDBEQSVQLQCTLOEDYRGTSTSVKTVLODERQSQALLIKADY 132  
DB 75 QDPQOQALQDLADANIDELAVALFFAAQEE-----AQVLDRLPLIAIMR----- 120  
QY 133 YEEERTCILRCV-LHLTYFODERHPYREYAVDCVKEKELVSKYRQGFELYKTEAPT 191  
DB 121 FHERHFLBELRLILQESFEVER-----EVTQALMDQWVA-----FVLEKXT 163  
QY 192 WETGNLWTERQ-----VSRRFV-----QCLREQS 216  
DB 164 GPLNNSSLFARKCKMSMEDIGWILLLEGEQIQKASIVGEVDROVWEAIEYORSLSQQQH 223  
QY 217 MLLEIFLYIAYFE---MABSDLVLTCKFKEQSGQTRHLVDENMDPPVDIRIGFS 273  
DB 224 ESLALICY--FFKGPYTSPEDLRYL-----LNRKLREFDGL--LHYIP 266  
QY 274 ALI--LVEGMDIESLHKCALDDBREH---QPAODGLIQDMDCLMTFFGDIPIHAPVLT 328  
DB 267 AMISSFVQHGSPERN--SYKARSLLHATVSTKDG-----QWTPIPF---HAAVIA 314  
QY 329 AMALLRHLNPEETSSVVRKIGTATIQNVFOYLTRLQSLASGNDCTTSTACWCYGL 388  
DB 315 LMLAVYSGMDIDGPTSPVQ---GVDFEKEABERTQMFTALDDG----- 356  
QY 389 LSFVLTSLHATLGNQODIITPACVTLADPSLPFLWG---TEPTSG-----LGII 436  
DB 357 LDFPLA---ICSGVNNEMADPARRELVALLKESAFANLSEPCAGFLKTLMLMENEIF 413  
QY 437 LDSVCGMPHLSP-----LQLRALVSGKSTA-----KVVYFLDMGSPFN 479  
DB 414 VESCIANMPDAVRKLIKSEEDMORLDQITALRDGLSSLRHGLVEARTLHESFLMIAIAF 473  
QY 480 ELYKHPADVISHEDGLM-----RRQTP-----KLIVPL-GGQTNLRIPQSTVGQ 524  
DB 474 ESRDAQEFWADDPGNLYGFLQWASKRQTVPRVSAFELLCSISGSEMAAARHFLTE 533  
QY 525 VMLDDRVLVWEVSY-SWTLFTCEIMLHV---STADVIOHCRVYK----- 571  
DB 534 ---EDKFLSKFKSTSMNWSQMFALQLYATRYTEKSPASQAVLRARXSEPADWSEPS 590  
QY 572 ---IDLVAHVSTDLSTADCLPITSRIYMLQELTTVISPPV-----DIVASCV 619  
DB 591 PVMLTLYRLMGLCKQSGSIREMMLQHP--FSVSVTLILCSGPITTHLRATVFTTLA 648  
QY 620 NCLTVLAARNPAKWTDLRHTGFLPFVAHPVSVLSQMSABGMNAGVGN-----LLM 672  
DB 649 ALMTDRTSNGNEMWL-----SLDQWISGAMSGPGLGKVPYVSNPLVW 692

QY 673 NSEDPQGEYGTI---AFILITTLVKGQUGSTQSGVPCWFLKEMLPYHKRRYN 728  
DB 693 HEQAFQKIGSEFPQANFVVLISLTPTSDSTREHSLP-----FPESIGS---SYR 743  
QY 729 SHGVREQIGCLILHAIHAIHNLCHERTDLHSHTSLOFLCI-----CSLAYTEAGQTVIN 783  
DB 744 MPGLIEPIYDFVMGQALSKRV-----PDIGERQTLTLTYNCLNPFYITCLRSFNENIVLYLS 798  
QY 784 IMGIGVDTI-----DMVMA--AQPRSDGAEQGGG----- 811  
DB 799 OPTISSDSLSLTSSLTYYIRLHPFARVAEWLFNEDVKAIFATQOQDASEVANASDSIL 858  
QY 812 -QLIKITVYKLA-----FSVTNNVIRLKPPSNVSPLEQALSQ-HGAHGNL--IA 857  
DB 859 IQCLVKSIEVNDMLDLQSTYFHLVRSIYKQAGSRINVANSSLSFEDSIANNLTITP 918  
QY 858 VLAKYIYHKDPALPRALQILKRLATVAPMSVYAC--IGN-----DAAA 900  
DB 919 ALSYICGAGHE-QLTVSMALLEKSSSRKLNKLSPEISWQSNKIVEYLSTEVDVDS 977  
QY 901 IRDAFLTRLQSKIEDMR-----IKWILEFLTYA--VETQGLIELFLN----- 942  
DB 978 VSRPLVSCMPLELELELGARSPGYIIRESILALINSLSMITTPRIANHLFGPSSVGT 1037  
QY 943 -LEVYKSDSGSKERSLGMWGLHVLLEIDS--QQQDRWCYPPLLH--RAAIAFLHAIWQ 997  
DB 1038 MLDISPN--LFSNGM-SLHHAITGFLQSYPDMABSNIIIPWVHLRMLAEVLKHLMS 1092  
QY 998 DRRDSAMLVLRK-----PKFW----- 1014  
DB 1093 SKIASAFTLEMRSRFLMSFASQPIIGPTPPDGLSVNDEEFWISSTSALEFLFR 1152  
QY 1015 -SYLYAVATEIRAAKIGSQTLQTEILSTLIGNSTTEGETILNPVFDLPFADLDISH 1212  
DB 1153 SYLYAVATEIRAAKIGSQTLQTEILSTLIGNSTTEGETILNPVFDLPFADLDISH 1212  
QY 1054 VV-----KSLDOS-----LKQTLKFKSLEKRFAYSGVYKSLAVHVA 1091  
DB 1213 KIPARPLVFDGIPDVCASQADBSLVNLAVALIQRKBELSSGQLRPODEEQF 1272  
QY 1092 ETESSCTSLLEYOMLVSAWMLIATTHADIMHLDTSVVRQL---FLVDLDGTVAL 1147  
DB 1273 SAEAGLT-----LPIRATNQ-----RLAPRIRYALASWIE-- 1305  
QY 1148 LVLPAVNGCRL-GSMKCTLL--ILRQWRBELGSVDEI---LGPLTEILSGVLO- 1197  
DB 1306 -LITMLACSEIEGGRKPTFIHISIQVLVPLAIAIEEDLPEAIEIARLAETLLIGLEFT 1364  
QY 1198 -----ADQQLMEKTYAKVPSAFTV--LQMKEMKVDIPOYSQVLNVNCELT 1242  
DB 1365 STSTDGARRSGDIIDEKHLQFOICIRGIFILASGNVQREFFYNGISYIRIAS-ADTV 1423  
QY 1243 QEEVIALFPDQTHSLALGATEDKDSMET--DQCSRSRHRQORGVCVLGLHAKELCEV 1300  
DB 1424 NQNL-----RHHSQOYIKTAGSTLIETDDDAVAGEACRVASALLELN--LATL 1471  
QY 1301 DEDGDSWL---QVTRRILPILPTLLTLEVLARKQN-----LHTEATYLLLTART 1350  
DB 1472 DRETDSTLAEILISQSNVLSLFLDAIRALPHELKATQANDTPALSYESLSLQRLQOT 1531  
QY 1351 QQCATNVAAGITQSI-----CLPILSVQOLS---TNGRAQTPSARSKSLDAPSWGV 1400  
DB 1532 KNGATQVLTGELFQAVBSRFLPADPDIGIGKCAISLTITDIDNPDAKRYKDYLG--SV 1589  
QY 1401 YRLMSILM-----EQLKTLRYNPLPEALDPGVGHQERTLOCLNARTVQSLACEEA 1453  
DB 1590 LRVLIISAVFSRGLNEQIMQOTRAFLENRQSWGIF--KRFAGIGVGTADHHHDLSDL 1647  
QY 1454 DHTVGFILQLSNPMK 1468  
DB 1648 AKCYMTLISATNPLE 1662

```
RESULT 30
US-10-437-963-122168
; Sequence 122168, Application US/10437963
; Publication No. US2004012343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 122168
; LENGTH: 2621
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_25121C.1.pep
US-10-437-963-122168

Query Match      1.9%; Score 169.5; DB 16; Length 2621;
Beet Local Similarity 18.3%; Pred. No. 0.00025;
Matches 350; Conservative 273; Mismatches 665; Indels 627; Gaps 86;

Qy 6 KITSVLSFCSRSRLM-----TILGSALELSQIAELMKMRLLLEGISYKRP 56
Db 481 QILNELACCRMLMALQKEMLELTATLTPEKARKEVEEQRHILCSNKRVLSNLDIEL 540
Qy 57 PPSPAEKVKANKDVASPLKEGLRISKFLGDEBSVOLQO-----YLQED----- 104
Db 541 SLASIKEMNGSNRCADL--EGBLRSTK--ENMERTIVELASCRNSLETLQNDMLLSAN 597
Qy 105 YRGTRDSKYTVLQD---EROSQALIKIADYYIEERFCILRCVLIHL----- 148
Db 598 SSFEKAEAIKLEEDMLCSNKEQGLLDLSETEK-----LHLSYAKHEHLESHAR 648
Qy 149 ---TYFQD-----ERHPYRVAVDCVCKLEKELVSKYRQOPEEL----- 184
Db 649 DMEYTFQGLTEQLLEENIYTSVDIYQITIKELYAKCNVVLGEGARNAHONNEACLDSE 708
Qy 185 -----YKTEAPTWETHGN-----LMTERQVSRWFVQCLREOSMLLEIFLYAYPEMA 232
Db 709 IIVENVERETTSPELIGHDDNORPLVAEND-----SC---NSTALQSL---KGHLKVA 756
Qy 233 PSDLIVLTMRKEGFGSRQTRNHLVDETMPFVDRIYFSALLIVEGMDESILHKCALD 292
Db 757 KGDRLDLOKLYER--ISSRSDGRVLSKLIQSFESK--GNQBDLGMSEBEN-----D 804
Qy 293 DRREIHQAPAOGLICODNDCLMTFGDI:PH-----HAPVLLAMALLHTLNPEETS 343
Db 805 NLRKLTQ---EMICRLVEKIKAMTSIAKTEBYVAELCNRIEISYKMSQSHAEIETHA 860
Qy 344 SVRKIGGTATQI--NVFOYLTRLQSLASGNDCTTSTACMCVYGLSLFVLTSLDELHTLG 402
Db 861 VLVAKMDFAGKLSYKQTIQVLQSVAVHODADNHNAGRLIDQAEILLQNVTERISITLE 920
Qy 403 NQO--DIITACEV-----LADPSLPELFWGTEPISGIG--IILDSV--CG-----MF 444
Db 921 KERTSLTDVLMVEVDTKLSALSKNALPSDIGSE--GLGSLATSSVECAALVONTQEKL 977
Qy 445 PHLSPLLQILRALVSGKST-----AKKYSFELDKMSFYNELKKKHPHYI 490
Db 978 EHAQDNNAKLAASLVELKTAHSDVOERSKAHAGIYKMYISLQELLP--NSLGN----- 1029
Qy 491 SHEDGTLWRQTP--KLTVPLGSGQTNLRIPQGTGQVMDLDRAYLVWRMEYSVSWTLPTCE 549
```

```
Db 1030 PDESGVEYNABEPIEALFSQYGDIVERHL-----KSLHERQYLSKNTDLESLSKCE 1083
Qy 550 IEMLIHAVSTADYVQHOCQKVPITIDLVKXITSDLSIAD-----CLPIRSRIYMLLQRL 604
Db 1084 ETEAL-----SSSLTKSNMDFSLNLEELKSVSTSRIEADQLHGRCLIAEKRVSRSTSHS 1139
Qy 605 TTVISPPVDVYASVCNCLTVLAARNPAKVTMDLRHTGFLPFVAPVPSLSQMSIAEGKNA 664
Db 1140 STVLS--SMEMSSKANHLITL-----LPCIEGVASYLE--EFEM-- 1177
Qy 665 GGYGNLLMNSBPQGEYGVTAFLRLITTLVKQGLSTQSGVPCVMFVLEKMLPSYHK 724
Db 1178 -----ABE-----IRSKICQESNIGQ--SSSEKMSVSLPVLIKETVLPF-- 1218
Qy 725 WRVYSHVRQIGCLLIELHALINLCHETDHSHTPSIQFLCISLATYEAQYVINYI 784
Db 1219 --PDLQGRIDQLSTLNIQL-----ETEPV 1241
Qy 785 MGIGVDITIDWMAAPRSDGAGGQGLIKTVKLAFSVTNNYIRLKPENNVSPLEQA 844
Db 1242 LRDSUTLQDSAL-----ETSRALQO-----KVPFELQSRQKLSYKEX 1280
Qy 845 LSGHGAHGNMLIA---VLAKYIYKHPALPRLAIOQLKR---LATVAPMSVYA----- 892
Db 1281 LSIIVAKGKGLIVQDLSKQTLLEK--SGELEKLAHELQSKDLSLIEAKIKSYADADRI 1339
Qy 893 -----CLGNDAALRDAP-----LTRLQSKIEEMRI-----KNNILFELT 927
Db 1340 EALESELSTYRNSKTALRDFLOKDSVLRQRIEVELDLDPENPHFDIVKELLSKMA 1399
Qy 928 VAVE--TOPGLIELEPLNLEVKDGSPGKEFSLGMMSCHAVALIELDSQOQD----- 976
Db 1400 VGASFYTP-----DGNKQSSVDGNSGCAIIDSINDQNSNSNGAREIK 1444
Qy 977 -RYWCPPLLRALAFIALHLMQDRDSAMLVLRTKPFWMENLTSPLFGLTSPPESTSEPS 1035
Db 1445 IKY---DELHR-----RFYE-----LAENHNMLEQSS 1467
Qy 1036 ILTECALIMKICIEIYVVVKGSLDQSLKDTLKKFSTIEKRPAYSGVYKLAIVAVATEG 1095
Db 1468 LVERNNLIQK-----NEEVLQOI--SIPOQRLPEDEBIA--WLG----- 1504
Qy 1096 SSCSTLLEYQMLVSAWMLLIATTHADIMHLDPSVVRROLFDLVDLGTALLVLPASVN 1155
Db 1505 ---NRLLEVEHERA-----LHKIEHLEBS--SEMILSDLESBK----- 1540
Qy 1156 CLRIGSKKCTLLILRLQMRRELGVSDEILGPLELIEGVLOADQOLMEKTKAVPSAFI 1215
Db 1541 --RISELSAEIVAV-----KAEKEFFSQSL 1563
Qy 1216 TVLOKMEKMYSDIPQYSQVLAVNCEITLOEEVIALFDQTRSHLSLGSATBEDSDMETDCS 1275
Db 1564 EKRFDFGLJSE-----KAVODEFVR--DWLRKDLA----- 1592
Qy 1276 RSRHRDQDQCVGLGLIAKELCEVEDG---DSWLOVTRRLPIPLTLTLEVSIRMK 1331
Db 1593 -----BLQEKLAKEBESKLYHDMEMEKIKMDLVADALQD--DSNTEIP 1635
Qy 1332 QNLHFEATLHLITLARTQOG--ATAVAGAGITQISICLPLISYQSLSTNGTAQTPSARK 1390
Db 1636 SGAGVGAIVAVCLGSLIRLLIDGYKTHLSESTVSSAEMETLSETKISKIDASTSERGMEK 1695
Qy 1391 SLDAPSWPQYV---RLMSLSMEQL-----LTIKRNFLPEALDPFYGVNOER 1433
Db 1696 EMALNTLTSGELEHTRNSLAVVEQDRBAVERKQTLTTELETLRQI--DQIQGDGADQMN 1753
Qy 1434 TLQCLNAVTRVQSLACLEADHTVGFILQSNFKEMHFLPQLMRDIOVNLGYLCOACT 1493
Db 1754 RYQSL-----MLELESMTKQ-----RDLQELTGGEBOEOKCT 1784
Qy 1494 SLHSRMLQHYLQNKNG-----DGLPSAVALQ-----RVORPPSAASAPSSSKOPA 1540
```

Db 1785 SL-----REKLANVAVRKGKGLVQHRDLSLKOTMEENMTMEIKLAKVERKQIHESL--ESERSSL 1839  
Qy 1541 ADTEASEQOAH-IVQVGLKLKLSKTLAALNHPDPVCOILLDGLAEVNFALSF 1539  
Db 1840 MGRLAENEKSHDHTQY-----LSRLNLSL--TVDIGREBDTDPITVENFSKCLDLQ 1892  
Qy 1600 TPTPDSVAPSPFGTLTATVNV-----ALNMLGELDKKKEPLTOAVGLSTQAGTRTKSL 1655  
Db 1893 NEVKSQKATEL--LLAEINVEHERADNLODELVKAELASESFRKONSVESAR----- 1944  
Qy 1656 MFTMENCFYLLISQAMRYLRDPVAHPRDKQRMQ-----ELSELSTLLSL 1702  
Db 1945 -----ADAVRHL-ERIMMOSQTRKQIDHLMELNSTSSQLREIFSEL 1986  
  
RESULT 31  
US-10-437-963-189782  
Sequence 189782, Application US/10437963  
Publication No. US20040123343A1  
GENERAL INFORMATION:  
APPLICANT: La Rosa, Thomas J.  
APPLICANT: Zhou, Yihua  
APPLICANT: Cao, Yongwei  
APPLICANT: Mu, Wei  
APPLICANT: Boukharov, Andrey A.  
APPLICANT: Barbazuk, Brad  
APPLICANT: Li, Ping  
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
FILE REFERENCE: 38-21(53221)B  
CURRENT APPLICATION NUMBER: US/10/437,963  
CURRENT FILING DATE: 2003-05-14  
NUMBER OF SEQ ID NOS: 204966  
SEQ ID NO 189782  
LENGTH: 2122  
TYPE: PR  
ORGANISM: Oryza sativa  
FEATURE:  
NAME/KEY: unsure  
LOCATION: (1) (2122)  
OTHER INFORMATION: unsure at all Xaa locations  
FEATURE:  
OTHER INFORMATION: Clone ID: PAT\_MRT4530\_86258C.1.dep  
US-10-437-963-189782  
  
Query Match 1.8%; Score 166; DB 16; Length 2122;  
Best Local Similarity 19.3%; Pred. No. 0.00036;  
Matches 344; Conservative 245; Mismatches 585; Indels 612; Gaps 87;  
  
Qy 166 VDKLEKELVSKYRQOFEEELYTEAPTWETHGNLMTROVSFWVQCLREOSMLEITFLY 225  
Db 739 IDKLEKCTELVSKYF-----NNDTEALNARILIN--DGNLTLLDDEL- 780  
Qy 226 YAYEMASDILLVLTKMKEGSGSRQTNRH-----VDETMDPFRDICYSAALIVBGM 281  
Db 781 --FRTSGKTIIF-----OKHLYIMVNCIAPFQISKTFVDEGFSAGV 823  
Qy 282 DIESLHKALDRRELHQFADGLICQ--DMDCMLTFGDIPIHRAVYLAMALRLHTLN 338  
Db 824 RVESLMLMA-----SICSLILGFCMLP-----LAHR-N 852  
Qy 339 PEETSSVVRKGTGTAIQNLNFOYLTRLQSLASGNDCTTSTACMC--VYGLSLVLTSL 396  
Db 853 KDVISSALKCTEGSL--VWQ--RLSASLSRNGN--SKLPKCMLSPTFGVF----- 898  
Qy 397 ELHTLNGOODIITDACEVLADPSLELFWGTEPTSGGLIIDSVCWMPHILSPLOLR 456  
Db 899 -LGSIVNCKTMSDTRFL-----PAYISLSLPSQ 928  
Qy 457 ALVSGSKTAKKYVSLDKMSFYNELYKRPKPHVDISHEDGTLMRQTPKLYPLGGQTMAR 516  
Db 929 DLMVPEH-----LHERFDGSTYDAIILH-----FILRSKMKLSYGKFMILS 969

Qy 517 IPQSTGVQVMD-----DRAYIVRWEYSYSSWTLFCEIEMLHVSTADVIQ 564  
Db 970 ILKG-VSILFVEDVSKSLFDLLR-----RNQVSG-----CESR--QIMSTHERQI 1015  
Qy 565 HQCRKPIIDLVHKV-ISTDISIADCLPITSRIYMLLQRLTIVISPPVDVIAVCVNCIT 623  
Db 1016 LCILLVWFVSNSANSSERS--EALLKVL-RIDVGAQDDPVVMP-----CVT 1062  
Qy 624 VLAARNPAKWTDLRHGTFLFVAHPVSSLSQMTISAGMNAAGYGNILNASEQOGEYGV 683  
Db 1063 ALQAVQP--VEFDLKTIDQEKVAFSLISMERTEITERNAARDAILIRINVA-----ST 1115  
Qy 684 TIAFLRLITTLVKQQLSGTSGQ-----GLVPC-----VMPVLKE 717  
Db 1116 AVKFIELIAQGDKKMSKRIKREDLNHDIKFNPDLLGVKPPASVLSLIDVLFKGD 1175  
Qy 718 MLPYVHKWRVNSHGVREIQGL-----ILELHAILN-----LCHETDLH--SSHTBSLQ 765  
Db 1176 VI-----QRTCLLOPLFQLSKLSLSDQMLIGIVQYKNGHDSAPENPDL 1220  
Qy 766 FLCISLAYTEAGTVININGIGVDTIDWMAAPRSDGAEQOGOLITVKALESV- 824  
Db 1221 NFWI-----BAQGLVLLIKDITDTLQ-----SGHQBELFCRDNILLINCIOAKDILG 1269  
Qy 825 TNNVIRLKPSPNVVSPLEQALSQHGAGNNILAVLAKIYHKHPALPRLAIQILKRLAT 884  
Db 1270 TRN-----HGFSILASLAK-----APQVVSSEIBDL- 1296  
Qy 885 VAPSVYACLDGDAARDAFLRLQSKIEDMRISKVMIIEFLT----- 927  
Db 1297 --FAIGD--AKQFLIKLVNVEHRRLTLWVYLLRTIGERKCLSTVMCLHS 1347  
Qy 928 -----VAVETOPGLIELPLINEVKDQSGSKESLQGM- SCHAVLELID 972  
Db 1348 LVGRISHSPHQALSLRAMPQEWEGYGLAVNI-----TNGYSYKLMFHCSKLQOET- 1399  
Qy 973 QOQDRYWCPRLLHRAALAFALHMODRRDSMLVLRKPKRWENLTSPLGTLSPSPSTS 1032  
Db 1400 RVHERQYVLPWLH-LAQOFLIFKLD-----TELIFPL--DSEEA 1436  
Qy 1033 EPSILFETCALIMIKIETIYV-----VKSLDQSLDKTLKKESEKRFAYW--SGV 1083  
Db 1437 ANSIGSGELMKVLCVAVNRDKTGISGDALKELRDSAN--TILKVIAWMCASYTF 1494  
Qy 1084 KSLAVHVAETEGSSCTSLLEYQMLVSAMRMILITATTHADIMHLDTSVVRQLFLDVIDG 1143  
Db 1495 KGI-----SRMLBH-----SRSVKRRKAL----- 1513  
Qy 1144 TKALLVPASVNCARLQSMKCTLLILLRQWK-RELGSVDEILGRLTEILEGVQADQOL 1202  
Db 1514 --GILCETAKNS-----LLOKQKRAKLHST--PAT--ALQVD-- 1548  
Qy 1203 MEKTKAKVSAF-ITVLQMKEMKVSDFPOYSQVLVNCETLOEBVIALFDOTRHSIALGS 1261  
Db 1549 --KSAPCFSELCKYKILEVNDREYDSVSVAIAAISLETIAKYPS--DMPATRKCLAK 1604  
Qy 1262 ATEDKSMETDSCSRSHRRDRDGVCVLGLHLELCEVDEDSQWLTQYTR----- 1313  
Db 1605 ITNHNSGDVAVTSRSRIY-TVGSLINVLG--SKALPOLPLIMKMLQVSHQVFCPSGK 1660  
Qy 1314 -----LPIPLTLTLEVSARMQNLHFTATYLLHLLTL-----AR 1349  
Db 1661 YAHSTKTDAVLSNOAIPILLSVLTVEVI--VKKLGEFVAPVLEEIIDLVLHPECASR 1718  
Qy 1350 TOGATAVAG--AGITOSICL-----PLLSVYOLSTNGTQOTPSARSKSIDASWPQV 1401  
Db 1719 NDEKLDKADAVRKLIDDKVVRMLSPILNLY--NGAIKCEBA----- 1760  
Qy 1402 RLSMSLMEQLIKTLRYNFLPEALDFGV--HOERTLOCLNA-----VTVOSLACLE 1451  
Db 1761 --SLIAFEMLSLST-----LVGAMDLNAVGTHTKYVEHCVALDLRQHLDSLKNIAIVE 1813

QY 1452 EADHTVGTLOSTNFMKEWHFHLPOLMRDIOVNLGYLCOACTSLHSRMLQHYLON--- 1508  
 Db 1814 QS--IIHATITMLTENTFR-PLFLRILEWABSEVDRSTSKSMDRAIVYKLVNSLA 1870  
 QY 1509 -----KNGDGLPS-----AVAQVQRPSPSAASAPSSSKOPADTE 1544  
 Db 1871 EKRRLGLVLPISVRWMPGWSLFTPYFKYLLGEGSVQVLSSEDDALLSSKOKKKALJEDAP 1930  
 QY 1545 ASEQQAHTVQ-YGLKILSKTLALRHPTPVQCLLDOSJLDLAEVYPLSTTPTTF 1603  
 Db 1931 VEQKKLGGPKLMNRALVLSLH--KCPLYNDKILDS-----NFOALKPIVSQF 1982  
 QY 1604 DSE-----VAPSGTLLATVNVVNLML---GELPKKCEPLTOAVGLSTOAGETRL 1651  
 Db 1993 VIEPEHESVPEASVDEVDLVLGCGQMAVTRSDVLMKPLNHEVGTGKHYASS--- 2039  
 QY 1652 KSLMFTMENCYLLISQAMRYLRDPAYHP-----RDKQRMKEISSLSTLSLSRYF 1706  
 Db 2040 -----FELDNGVLMKTRSDKVRPKMLGLKVVVRHWVQQLKEEYVVL----- 2081  
 QY 1707 RGAPSPATGVLPSPQGSTSLSKASPSQEPILQVQAFVRHMQ 1752  
 Db 2082 -----PBT--IP-----FLAELLEDVELPVKTLLAQEIVKEME 2111  
 RESULT 32  
 US-10-128-714-3246  
 ; Sequence 3246, Application US/10128714  
 ; Publication No. US20030119013A1  
 GENERAL INFORMATION:  
 ; APPLICANT: Jiang, Bo  
 ; APPLICANT: Hu, Wenqi  
 ; APPLICANT: Tishkoff, Daniel  
 ; APPLICANT: Zamudio, Carlos  
 ; APPLICANT: Broshkin, Alexey M  
 ; APPLICANT: Lemieux, Sebastien M  
 TITLE OF INVENTION: Identification of Essential Genes in *Aspergillus fumigatus* and  
 TITLE OF INVENTION: Methods of Use  
 ; FILE REFERENCE: 10182-018-999  
 ; CURRENT FILING DATE: US/10/128, 714  
 ; PRIOR FILING DATE: 2002-04-23  
 ; PRIOR APPLICATION NUMBER: US 60/285,697  
 ; PRIOR FILING DATE: 2001-04-23  
 ; PRIOR APPLICATION NUMBER: US 60/287,066  
 ; PRIOR FILING DATE: 2001-04-27  
 ; PRIOR APPLICATION NUMBER: US 60/295,890  
 ; PRIOR FILING DATE: 2001-06-05  
 ; PRIOR APPLICATION NUMBER: US 60/303,899  
 ; PRIOR FILING DATE: 2001-07-09  
 ; PRIOR APPLICATION NUMBER: US 60/316,362  
 ; PRIOR FILING DATE: 2001-08-31  
 ; NUMBER OF SEQ ID NOS: 8603  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 3246  
 ; LENGTH: 1545  
 ; TYPE: PRT  
 ; ORGANISM: *Aspergillus fumigatus*  
 US-10-128-714-3246  
 Query Match 1.8%; Score 165; DB 14; Length 1545;  
 Best Local Similarity 17.8%; Pred. No. 0.00025;  
 Matches 304; Conservative 259; Mismatches 628; Indels 516; Gaps 76;  
 QY 96 LLOC--YLOEDYNGTRDSVKTVLQDERQSQALILKADYYEERTCILR-----C- 143  
 Db 10 LLECIARLIGSFEVEREVTOALMOD-----MVAFVVEIKTGPRLNRSNLFARKCM 59  
 QY 144 -----VHLTLTPQDERHPRVREYADCVDLKELEKLVSKYQOQEBEL-----YKTEA 189  
 Db 60 KSMEDISGWLILLBEOIQKASIVGQVEDRVMEALIEYOR--SLIQOQHESIGAILCYFPKG 118  
 QY 190 PTWETHGNLMTROVSRWFVQCLREQSMLEIIFLYUYAVFEMAPSDLLVLTQMKEGQFG 249

Db 119 P-----YTPSPEDAVLNLRKLKER-FDGLLHY-----IPAMISFVQHGSP 160  
 QY 250 SRQT-----NRHU--VDETMDPFVDRIGYFSALIL-----VEGMDIE 284  
 Db 161 ERSNSYKARSLSHLAVSTKDGQWTIPTFHAIVIALMLAVYSGMDIDGPTSPVQGVDFE 220  
 QY 285 SLHICALDDREHLQFQOD--GL-----ICQDMCLMTGTGDIPIHNAFVLLAMVLAHTL 337  
 Db 221 K-----EABERTQMWTMLDGGDLDFLALICG-----V 249  
 QY 338 NPEETSSVVRKIGGTAIQLNVFOYLTRLQSLASGNDCTTSTACMCVYGLSFEVLTSLE 397  
 Db 250 NNEEWADPAR-----ELVALLKESAPANLESPP--CAGFLKTLTMENTLE 293  
 QY 398 LHTIGNOODIITDACEVLADPSLEPELFWGTPTSGIGI IIDSVCMPHLLSPILLQLLRA 457  
 Db 294 IFVESCJANMPDAVRKIKSEDMQR-----LDQITLALDGLSS--SLHRG 336  
 QY 458 LVSGKSTAKKYVSFLDMSFYNELYKHKPHDVISHEGTLW-----BRQTP----- 503  
 Db 337 LVEART--HLESTLMIATAFESRPAOEFMAADPDGNLYGFLQMASKQVTPRVSAFC 393  
 QY 504 KLYPL-GGQTNLRIPQGTGQVWLDRAVLVEMEYS-SWTLFTEIEMLLHV----- 557  
 Db 394 ELLCSISGGEENAAAHRLFTE--EDKFLSKFKSTSNMWSQMFALQVATRVTEKP 450  
 QY 558 STDVIOHCQRVP-----IIDVHKVISTDLADCLLPTSRIMYLOR 603  
 Db 451 SASQAVLRARKSEPADMSPEPSVWMLTCYRLMGLHCKGSGSTREWWLQHP--FSVAVST 508  
 QY 604 LTTVISPPV-----DVASCVNCLTVLAARNPAKWTDLRHGTFLPFVAPVPSLSQMI 657  
 Db 509 LTLTCSGPITPHLRATVFTTLALMTDRSHNGNEMWL-----SLDQWI 552  
 QY 658 SAEGBNAGRYGN-----LANSBOPQGEYVTT--AFRLITTLVKQGLSGTOSQG 706  
 Db 553 SGGAMSGPGKGVPIVSNPLVWHEQOAFKIGSGFQDANFVVLISLTPTSDBSTEYHL 612  
 QY 707 LVPCEWVLKEMLPYHKMYRNSHGVRBOIGCILELHAILMLCHETDLSHTSLOF 766  
 Db 613 SLP-----FPESIGS--SYRMPGIEPIYDFPMGQALSKV--PDLGERQTRLITY 658  
 QY 767 LCT-----CSLAYTEAGQVYINIMIGVDTI-----DMW 796  
 Db 659 NCLNFVITTCRSFNMENIVTVLSQPTISSDSLSKTSITTYIRLHPARVAEMLFNEDEVIK 718  
 QY 797 A--AQPSDABEQGOG-----QLIKTYKLA-----FSTYNNVIRLKPFSNVV 838  
 Db 719 AIFATMOQDSEVANAAADSILIQCLVKSIEVMDMLDQSTYFHVIRPSIKSQAGGSRI 778  
 QY 839 SPLEQALSQ-HGAGNNL--IAVLAKYIYKHDPALPRLAIOLEKRLATVAPMSVYAC-- 893  
 Db 779 NVANSSISSEFDSILNLTITPALSLYCGAGH-QLTYSMALLEKSSSKRLKLSPE 837  
 QY 894 LSN-----DAAIRDAFLTRLOSKIEDMR-----IKWILBELVTA 929  
 Db 838 ISNWOSNKIWEVLSTEVDVDSVSRPLVSGMCPLELRELEIGARSPEGIIRRESLALLNSC 897  
 QY 930 --VETOPGLIEFLN-----LEVKDSGDSKESKSLGMWGLHVALELDS--QQQDR 977  
 Db 898 LSNITRPFITLHFLFGSSVGTWLDISPN--LPSNGM-SLHHAITIGFLOSYPDAMEG 952  
 QY 978 WYCPPLH--RAAIAELHALMODRDSAMLVLRK-----ENLTSPLGTL--SPSET 1031  
 Db 953 NILPMVHILRRMLLEVILKILMSKINSAFTLLEMRSRFLWSFASQPIIGDPTPDGLS 1012  
 QY 1011 ---PKFW-----EUTSPPLGTL--SPSET 1031  
 Db 1013 VMDSEFWISESTALAEFLFRSYLYAVATEIRAAKLGSOQLQTEIILSLGNSSTET 1072  
 QY 1032 SEPSILETCAILIKKICLEIYVV-----GSLSDQS-----LKOITLKK 1069  
 Db 1073 GETIILNPTVFDLPDPADLDLSHKIPAPRLVFLDGIQFDVCAKSQADESLVTLNLAEVAL 1132

Qy	1070	FSIEKRFYAWGGYKSLAVHVAETBEGSSCTSLLEYQVMSAKRMILLIATTHADIMHULTD	1129
Db	1133	IQVRBELLSSGODRPODEQFSABEGLT-----LFIATQVA-----	1171
Qy	1130	SVBRRL-----FLVDLGTCKALLVPASVNCRL-GSMCKTLL--ILLRQWKELGSDV	1182
Db	1172	-----RQIMFNRYLALRSWTE---LITTMLAGSEIGBKPFPIIHSIDULPKLEAIEE	1222
Qy	1183	EI-----LGPTEILEGVLO-----ADQOLMEKTKAKVSAFTIV--LOM	1220
Db	1225	DLPEAIELIARIAETLIGLEFSTSTGDARRSGDIIIDBKTLQDFQICIRGIFLASGNQL	1284
Qy	1221	KEMKVSODIPOYSQVLVANCETLQGEVIALPQTRHSLALGSATBEXDSEMT--DDCSRSR	1278
Db	1285	RETFPNISQÖITISRIAS--ADTVNÖML-----RHSQÖVIKTAGSTLIEFTCDAYAGQ	1336
Qy	1279	HRDQRGVCVGLHMLAKELCEVDEBGDSWL-----QVTRRLPLIPTLLTLEVSIMKON-	1333
Db	1337	EACRSVALLFLNLT-----LATIDRETDSTLAEILISQSVNLSIFLDAIRALPHEILKNTQAN	1391
Qy	1334	-----LHTEATIMHLLTLARQOQATVAAGITQISCLPLSYVQLSTNGTQOTPEAS	1388
Db	1392	DTPALLSYBESLSTLQRLCQTKONATVULKTGIFQAVARESBFLPAADPLGIDIDNPAL	1451
Qy	1389	RKSLDAPSPGCVYRLSMSLM-----BQLTKTLRYNLFPEALDPFGVGHQERTLQCLNAV	1441
Db	1452	RKYVDLUG--SVLARITISAVFSRGLAHNQIMQOTPAFLAENRQSWGIF--KRFPAKIGV	1507
Qy	1442	RTVQSLACLAEADHTVGFTILOLSNPMK	1468
Db	1508	GTADHHDALSDLAKCYMTLISATYNLE	1534

```

RESULT 33
US-10-437-963-114113
; Sequence 114113, Application US/10437963
; Publication No. US20040123343A1
GENERAL INFORMATION:
APPLICANT: La Rosa, Thomas J.
APPLICANT: Kovalic, David K.
APPLICANT: Zhou, Yihua
APPLICANT: Cao, Yongwei
APPLICANT: Wu, Wei
APPLICANT: Boukharov, Andrey A.
APPLICANT: Bardazuk, Brad
APPLICANT: Li, Ping
TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
FILE REFERENCE: 38-21(53221)B
CURRENT FILING DATE: 2003-05-14
NUMBER OF SEQ ID NOS: 204966
SEQ ID NO 114113
LENGTH: 2462
TYPE: PRT
ORGANISM: Oryza sativa
FEATURE:
OTHER INFORMATION: Clone ID: PAT_MRT4530_17836C.1.pep
US-10-437-963-114113

```

Query Match	1.8%	Score 160.5;	DB 16;	Length 2462;
Best Local Similarity	19.1%;	Pred. No. 0.0015;		
Matches	369;	Conservative 259;	Mismatches 609;	Indels 695; Gaps 106;

  

QY	25	LGRSLRLRLSQTLEA-----	ELNKKRRLLLEGLSYKKPPSPSSAEKYKANK	69
DB	810	LGVEASNLRLSHVLPFLCDALSVGNNTLYKYQDHTIRKLISKPNQEGYSPS-----		860
QY	70	DVASL-----KELGRI-----SKFLGDBEGSVQLQC-----YL-----QEDIRGTDSYKT		114
DB	861	--FSPILVCVQCKCRRLDSEASVYKLEHKSRTISLYVCNTVVLIMQSVQDVLPLDLVGT		918

QY	115	VLODEROSQALIKIADYYEERCT----	IAQVHLIYYFQDERRPAYEVADCYDKL	165
Db	919	ILINER-----IGFSSSEINSRICEFMBPMLMILHISISD-----		956
QY	170	EKELVSKYRQOFBELYLTETAPMTWETHONIMTEROVSMFOCUBOSML----	ELIFLY 225	
Db	957	-----QSSSLFSTLEHSSSVYGNLSIC--SVRTIEBMSNQPTNLPDVAISFVLY		1005
QY	226	YAEFEMAPSDILV-----LTKMFEQGFSGKQTRNHLVDETMDFVRI		269
Db	1006	SVI---CAPPDVISFPKLLHVLKTH--FPSSL-----PFLSSVFLQHDYLAKY		1051
QY	270	-----GYFSAIILVEG--MDIESLH-----	KCAL-----	291
Db	1052	ASYCPDMFFSSILROIKNLIDVDSYNIYEDKWKHSTCSBASIVSTPLNTPFCALPSVL		1111
QY	292	-----DDREHLQFPADGLICQDMCJMLTFEDI-----HHAFLVLMALLRHPLN-		338
Db	1112	SLAFSAPEITKAHPLODELV---HLQAKISSEPSLETTLRVVLFWIS--HHLISS		1165
QY	339	-PEETSSVVRKIGGTAIQL--NVFOYLTRLOSILASGANDCTTSTACQVYGLSFPV--		392
Db	1166	YTVACSDIADLGVCVCSLIDSIPIERI--RVLTADTANSKSVARYPVPCJNGIYESVQH		1224
QY	393	-LTSLELH-TYGNQODIITDACEV-----LADPSLPFLF-----		424
Db	1225	PIIGSLISCSISNFODIADGSEVEYKEDFASFSKEXHLADSFPVNLISMLYGLVLAGS		1284
QY	425	WGTEPTSGLGIILDSVCGMFPHILSPILLOLRALVSGKSTAKKYYSTFLDKMSFNEIYKH		484
Db	1285	FGANYSNNDGOSTLESGPPKLLERILLLFKKFEICMEKRNGLFLPNFYERTLAKF		1344
QY	485	KPHDVISHEDGTLWRROPTPKLLYVLGGQTMRLPQGTVGQVMLDDRAUYLVMEYS----	Y 540	
Db	1345	-----VSPRILE-----LANMNSTEGRF		1364
QY	541	SSWT-----LPTCEI--EMLHVYSTAD-----VIOHC-----QRYKPI		572
Db	1365	SSSPAPAYAPAFCCLYADIADIFEMLYDYLOQIDQSGPCRLMGLEHNCADIATIQOYNI		1424
QY	573	IDVHKVISTDLSAD--CLLPYISRI-----YMLQRL--TYVISPPVDVIAS		617
Db	1425	I--LHPATKMLERPADICLLKMLRLIHHTETISAVRNTDYIALHMMUSTWANTPDIHLH		1482
QY	618	CVNCITVLAABNPAKWTDLRHTGFLPFVAPHVPSLSQMSIABEGMASGYGNILMNSQEP		677
Db	1483	CM-----FPSKYKAAIQLLLGANGMHRRLFOKLLTD-----		1515
QY	678	QGEYGVTTIAFLRLITTVKVGQSGSTOGSLVPCVMFYLKEMLPBYHKM--RYNSHGVBEQI		736
Db	1516	-----LIKDOTSVWQ--VVGSDSNASWTHEDSFIL--LLPALSYIEHHSGNRQ--		1561
QY	737	GCLLELHIALINLCHETDLSHSTPBLQFCTCSLAVTEAGQVIVIMIGVDTIDMM		796
Db	1562	GSINNE-----LAPEKESKRVANNAKLFTISIL-----VATLUXIF		1596
QY	797	AAOPRSDGAEQOGOLIKTVKLAFSYTNVNLRLKPPSNVVSLEQA-----LSQGHAN		851
Db	1597	RNFPRDQIILSSBELNVTVC--LEKALIKRIVELS--SEVQSHNDQKPIPLNQ-----		1649
QY	852	GNNILAVLAKYIYHKHPDAPLRALIQLLKRLATVAP-----MSVYA		892
Db	1650	-----LIRSSILHRFSDP-----VVIKAIKCIIVVLSGKFPADIELELIGHSHFVSTIT		1700
QY	893	CLG-----NMAAIRDAFLRLQSKIEDM-----RIKXIMLEFLVAV		930
Db	1701	CSGVSECPACNPTGGLLOPAPSLTKSVDSAPAK--ENKQODCIPEKKVEIIRLRLKVLV		1758
QY	931	ETQ-----PGLI-----ELFLNTEVKGSDSKFSLGMMSCHLAVLELIDSQ-----QQ		975
Db	1759	DIKSRQHNSSILDSREBREGFLLLSVYGATISETDLEI-----LHMNEIESSBEKALITDV		1813
QY	976	DRWYCPBLHRAAIFLHALMOD-----RRDSAMLVLRTKPKFMENTSLPFLGTLSPP		1028

```

Db 1814 DLM-----GKAAVFRRELKLEPSASDTHGMENAEISDRRSLFREMI--PI-----D 1860
Qy 1029 SETSEPSILETC-----ALIMKIICT-----EYIVVAKSLDQ-SLKTQ--LKRSIE 1073
Db 1861 SKLCVMVTLQPCYRRSSRASIFSLQLOQDNFGDI FKATSSQMAVRYDVFILHSIH 1920
Qy 1074 KRFAVMSGVKVS-----LAVHVA-----ETEGSSCTSLLEYOMLVSAW 1111
Db 1921 ---TLMKYIIEPVERSLGLAITVLSISSPEDELRKIGYSLGTFFKSLASQSKETW 1977
Qy 1112 RMLLIATTHADIMELTD-----SVRRQDFLDVLDGT-----KALLVPASVN--C 1156
Db 1978 QLOQLLTLYLQNGISEWQORIPSVIAIFAAEASLTLLDSHTQPATISKFLMHSASVNLQ 2037
Qy 1157 LRLSMKCTLLILLRQKRELGSVDLGLPTEILEVLOAQDQMLEKTAQKPSAFIT 1216
Db 2038 LYASGN-----LADDAKIYKR--GGV-----LELALSYGSSAVSDSETKLLT--LQ 2079
Qy 1217 VLQMKEMKVSIDIPOY-----SOLVNVCELTQOEVIATLPDQTRSLA--LGSATEDKDSM 1269
Db 2080 VLK-KCVTLPLVAHLHTKDSGLMLMISSVISSHVEGL--DSVKNBSYSTVIGSALB----- 2132
Qy 1270 ETDSCSRSRHRDQBDGVCVLGLHLAKELCEVDEDEGDSWLYQTR----- 1312
Db 2133 VVNDLTSR-----LIAE-----WLGETALQSLRISKVLYVLED 2168
Qy 1313 ---RPIPLTLLTLEVSLR--MKQNL--HFTETHTLLLTARTQGAATVAVAG 1361
Db 2169 MKLKGANVPLLTSLVNLVASTMRLSMKRKIYQPHFS--LSLGHKLCCTI-----GG 2219
Qy 1362 ITGSICTPL--LSVQOLSTNGTAQTPSARSKSLDAP--SW--PGVYLSMLMEOLKLT 1414
Db 2220 ISRIETVGLAQGLDVIAMNGPLPVLSMDKSMRTATVSNATSNIFNL--CDEGRSVLK- 2277
Qy 1415 LRYNPLPEALDFVGVHDE-----RTLOCLNAVRTVQSLACLEADHTVGFILQISN 1465
Db 2278 ---MP-----HEEPLKNECLLSKMLRWLASIILKISCIS----- 2310
Qy 1466 FMKEMHPLPOLMEDIONLGVLOACTSLHSRKM-----LOHYIQ 1507
Db 2311 ---HEKGDPLTDAN--NFSL--BSFLNTYDEKVEVTSKSHSDEALAIILLYLQHLK 2363
Qy 1508 NKNGDGLPSAVA-----QVORPPSAAASAPSSSKOPADTEASEQALHTVQ---Y 1556
Db 2364 -MNRDPLPSVVAALCLLLDRSNKQSGNHVYSPRHYQPMKDPAMHNRBAHLEBEQACQ 2422
Qy 1557 GLKLKLSKTLAA 1568
Db 2423 SLVWFNSFSFA 2434

```

```

RESULT 34
US-10-408-765A-354
; Sequence 354, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Marnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; TITLE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
; FILE REFERENCE: 660088.465
; CURRENT APPLICATION NUMBER: US/10/408.765A
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 354
; LENGTH: 3859

```

```

; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-408-765A-354
Query Match 1.7%; Score 157; DB 16; Length 3859;
Best Local Similarity 17.3%; Pred. No. 0.0067;
Matches 335; Conservative 288; Mismatches 655; Indels 664; Gaps 90;

Qy 54 YKPPSSAKKVA-----NKDVASPLKEGLRI-SKELG-----LDEGQVQLQCTVQ- 102
Db 862 YDHIOFVRAELMQLMWTLEKRNPDGSIHVAHYVKGFGSSNRKMLKESQKLVHVVTEVG 921
Qy 103 ---EDYGRTRDSKTVLQDEROSQALILKIAQ---YYEERTCILECVLHLLYFQDER 155
Db 922 PSTIVERSDKASIQLEMEKAIETALDCLKSANTEPYRRQAMVEIKFLVAMMSLEDNK 981
Qy 156 HP-YR-VEYADCVDKLEKELVSKYRQOFEELYKT-EAPYETGNLNTERQVSRWFVQCL 212
Db 982 HALYQLAHNPFTEKTPNVYIISR-----YKAQTPARKTEQALTGAFMSA-VIKDL 1034
Qy 213 REGSM-LLEIIFLYAAFEWA-----PSDLVLVTMKFKQSGSRQTNR 255
Db 1035 RPSALPFVASLIRHYTVVAVAQCGPFLPCYOVGSPS---TAMFHSBNSK--- 1085
Qy 256 HLYDETMDPFVDRIGYFSALILVEGMDIESLHKCALDRRELHQFADGHLICQMDCLML 315
Db 1086 ---GMDP-----LVLDALAI---CMAYEEREL-----CKIGEVALA 1116
Qy 316 TFGDIPHHAPVLLAMALLRHTLNPEETSSVVRKIGGTALQNLNFOYTRLLQSLASGND 375
Db 1117 VIFPV-----ASIIISKERACQLPLFSYIVERL----- 1145
Qy 376 CTTSACMCYV-----GLLS--FVLTSLEI--HTLNGQODIIDRACEVLDPSLPEL 423
Db 1146 ---CACCEQAWAKLGGVSIKFERLPVTVWLNQOQTFKALLVMDLT----- 1195
Qy 424 FWGTEPTSGIILDSVCGMFPHLSPLQLLRALVSGKSTAKVVSFLDKMSFYNELYK 483
Db 1196 ---GEVSNGAVNAKXTLLEQLMRCATPLDEBERAEIVAAQER-SFH----- 1239
Qy 484 HKRHDIYSH--EDGTLRROTPKLILPLGGQTNRI-----PGTVQVWLDRAIYVRW 536
Db 1240 HVTHTDLREVTSFNSVTRKQAMSLQVLAQVTKGSVIVINEPHKEVLAQDVPVPRKGLLRH 1299
Qy 537 EYSYSN-----TLFTCEIEMLVHVSADVIQHCOR-----VKPI 572
Db 1300 QPANAQIGLMEGNTFCTTLPRLFTMDLVNVEKHVFTELLNCEAEDSALTLPCKSL 1359
Qy 573 IDLVHKVISTDLSIADCLPITSR---IYMLQRLTVISPPVDVIASCVNCLTVLAARN 629
Db 1360 PSIVPLRIALMLNLAACNVLPOGRREKIIALFALNSTNSSELQAGACAKRKFV-----E 1414
Qy 630 PAKVMTDLRHTGELPFAHVSLSLSQMSAEGNAGYGNLWNSBEPQOEYGTIAFLR 689
Db 1415 GATIEVDQHTHMRPL-----WMLGVRSLTN-----VVRRLT 1449
Qy 690 LITLVYQGLSTQSQGLVPCVMEFLKEMPSYHKRMYNSHGVBEQIGCLILEI----- 744
Db 1450 SVTRLPFNSFNDK-----PCDQWQHARKN-----MEVVVITHK 1483
Qy 745 ---HALINLCHETDLSHTPSLQF--LCISLAYTEAGQTVININGIGVDTIMV 795
Db 1484 GGORSDGENISISRCRPL-----SPFCQFEEMKICS-----AIIINLPHL----- 1523
Qy 796 MAQOPRSDAGEGGQGLIKTYKLAFSYNNVNIKLKPSNVVSPLQALSQGAHGNL 855
Db 1524 IPAP-----QTLVAPLLEVVMKTERAMLEAGSPREPLKFLTRHPSQVEL 1572
Qy 856 IAVLAKIYHKHPALPRLAIQLKRLATVAPMSVYACLANDDAARIDA-----FLTR 908
Db 1573 FMEEATL---NDPQSRMMSFLKH-----KQARPLRDVLAANPNRRTYL 1614
Qy 909 L-----QSKIEDMRIRKVI--LEFLTVAVET----- 932

```



```

Db      1615 LLPGAGGTAVRGPSTSTMRDLDFQAIKIISIVKNDSDSLASQHSIVSGLRRVWSE 1674
Qy      933 -----OPGLIE-LFLNLVYKDGSDGSEKESL-----GMSCLHAUL-E 968
Db      1675 NFOERHRKEMMAATWMEKPEKLAAYCLLVYCKKNYDIELLFOLLPAFGRFLCNMTFLIKE 1734
Qy      969 LIDSQOQRWYMCPRLLHRAIAFLAALMODRRDSAMLVLRTPKRWENLTSPLF----- 1022
Db      1735 YMEETIPKQY---SIAQKALFFRVDFDNPFGDEL-----KAKVLQIHLNPAFLYSPEK 1787
Qy      1023 -----GLSPSETSEPSILETCALIMKIICLEIYVYVXGSLDOSIKDTLKKEISIEKRPAY 1078
Db      1788 GGBGQLLGPENPEGNESITSVFTKYLDPE-----KQADMLDLIRIYLQ-----Y 1835
Qy      1079 WSGVYKSLAVHVAETEGSSCTSLLEYQMLVSAMRLIIA-----TTADINH-LTD 1129
Db      1836 ATLLVEHAPHNHIDNNKRNRSKL--RRLMTFAMPCLSKACVDPACKYSGHLLAHITAK 1893
Qy      1130 SVVRRLQLFDLYDG-----TKAL-----LVPAVNCIRLGSMTCTLLILLRQWK 1175
Db      1894 FAIHKKIVLQVPHSLIKAHAMEARAIVRQAMAILTPPAVPAEMEDGHQ-----MLTMT 1946
Qy      1176 RELGSVDE--ILGPLTEILEGLVQ-----ADQOLMEKTKAKYFSAFIT----- 1216
Db      1947 RKI-IVEBGHTVPOLVHILHLIVQHFKYVYVVRHNLVQHMVSAMQRLGFTPSVTIEQRRL 2005
Qy      1217 -----VLQMKEMKVSQDIPOYSQVLVNCETLQEBVIALPDQTHSLALSGATEDK--- 1266
Db      2006 AVDLSEVVIKWEIQRIDKQOPSDMDPN---SSGEGVSVSSSIRKGLSVDSAQGVKFR 2062
Qy      1267 -----DMEITDDCSRSHRQORQGVCLGLHLAKELCEVND----- 1303
Db      2063 TATGAISAVFGRSQSLPGADSLALAPIDKQHTDTVNVF-LIRVAQVANDNTVAGSPBEV 2121
Qy      1304 -----GDSWLQVTRRLPIILFTLLTTEVSLRMKO-NLHTEATILHLLTLA 1348
Db      2122 LSRRCVNLKTLARLDMWPKSELKQWFDKLMTEQNGVYVNGIICGLEVLSTLLVL 2181
Qy      1349 RT-----QOGATVAVAGITQISICLPLLSVYQULSTNGTAQTPSASRKSIDLAPSWP 1398
Db      2182 QSPALISFPKPLQRGIAACMTGNTK----- 2207
Qy      1399 GVVRLSMELMQLKTLAYNPLPEALDFVGNHQBERTLOCINAVRTVOGLACLEADHTVG 1458
Db      2208 -VLRAVHSLRLMSI--FTEPSTSSVASKYER--LECLYA-----AVG 2247
Qy      1459 FILQ--LSNFMKEWHFHLPOLMRDIQV-----NLGYLQCACTSLHS-RKMLQHYTON 1508
Db      2248 KVIYEGLTNYEKATNANPSQLFTLMIKSCSNNPSYIIDRLISVFMASLQKMAVEHLN- 2306
Qy      1509 KNGDGLPSAVAVQVRPPSASABDSSKQPAADTEASEQOALHTVOYGLLKILSKTLAA 1568
Db      2307 -----FOAASGS-----TEA-----TSGSELVMLSLBLVK 2332
Qy      1569 LRHFTPDVCOILLDSLDLAENPLFALSFTTPTPTDSEVAPSGFGLLTVNVALMMLCEL 1628
Db      2333 TR-----LAVVSMEMRK-NFIQAIL--TSLIKSPDAKILRAVVKI---VEW 2374
Qy      1629 DKKEPLOTAVGLSTQAGSTRLL--KSLIMFTMENCFTLLISQARVYLDRAVHPDRDQ 1686
Db      2375 VKNSPMA-----ANQPTLRKXSILLVKN-----MTYI-----EKR 2406
Qy      1687 MKQELSELSTLLSSLSRYFR 1708
Db      2407 FPEDL-ELNAQFLDLVNVYVR 2426

```

RESULT 35  
 US-10-424-599-256710  
 ; Sequence 256710, Application US/10424599  
 ; Publication No. US20040031072A1  
 ; GENERAL INFORMATION:

```

; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 256710
; LENGTH: 2627
; TYPE: PR
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1) (2627)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_73832C.1.pep
US-10-424-599-256710

```

Query Match 1.7%, Score 156; DB 15; Length 2627;  
 Best Local Similarity 17.0%; Pred. No. 0.0042;  
 Matches 326; Conservative 266; Mismatches 645; Indels 680; Gaps 83;

```

Qy      44 WRLLGLSTYKPPSPSAEKYKA-----NKDVASPLKELGLRISKPLGLDEQ 92
Db      392 WAPFESPPFKSSASSILRIRIELQENLCILHFRNVVISAMNDL-----MEISBE 444
Qy      93 SVOLLOCYLQEDYGTDSVTVAQLQDERQOALIKIADVYEBERTCILRCVLLHLLTFQ 152
Db      445 VILHRSFCR---KQQLDKQNSDFVDGTGRTALALMGAVSCT--AHMCTVGANPSLMEVL 499
Qy      153 DERHRYREYADCYDKLEKE---LVSKYRQOFELKYTEAPTEWGNLMTEROVSRMF 208
Db      500 DAVONPLVKSDDCIGMSKKAWSIIGALASFRLVSN-----SNGADETGK-----F 549
Qy      209 VOCIREQSMLEIIFLYAIFEMAPSDLLVTKMFKQEGFSRQTRNLVDETMDPVDYR 268
Db      550 LSLAKRYKSPQVLFVAVAGYLE-----FKGSLLEDVAVRYIHPLEBEKTDADA 597
Qy      269 IGYFSALIVGMDIESLHKALDREHLHQFADGILICODMCLMTFGDI PHHAPVL 328
Db      598 VATFS-----DNLHSDXERISTIKLCHYKP--L 626
Qy      329 AMALLRHLNPEETSV---VRKIGTAIQ--LNVFOYLRILQSLASGNDCTTSTACM 383
Db      627 GW-----ENSSVDQPAKRRKTEVSPTLNV-----ECTENNALL 660
Qy      384 CVYGLSFLVLSLHLYTGNQODIIDTACEVLADPSLELFWGTPTSGLGIIIDSYCGM 443
Db      661 L--LSTIETTPISISSRSIQLFISKIOMELASGRIPNVY-----VPLVINGLFGI 709
Qy      444 -----FPHLSPLQLLALVALVSGKSTAKKVSFL---DKMSVYNLYKHKHPDVASHDGT 496
Db      710 LNNRFSYLMNVLCEIAVLIS-----LHFLVWVSLVAVYERCO-----T 749
Qy      497 LMRQOTPLVPLAGGQTNLRIPOGTGQVMDADDAVYLRWEYSYSWTLFTCEI-EMLLH 555
Db      750 IF--DTBNLH-----GSVNGALFDQAGLV-----DCEFLFYVHNSDSGPS 789
Qy      556 VVSTADVIQHCQVKPIID-----LVHNVISRD-----LSTADCLLPITSRYMLQLRLT 606
Db      790 VTIALLLQALQIKPIVIEPRSRQIFLFLKGLPDLVSVGLFDSHACKKEWAKILIKE 849
Qy      575 -----LVHNVISRD-----LSTADCLLPITSRYMLQLRLT 606
Db      850 WNLILKLMKNPKSFYCGQFLKDVILQHRLLBENDTEIQMRVLDCLBELT-----TW 899
Qy      607 VISPPVVIASC-----VNCLFTVAARNPAKWTDRHRTGFLPFVA--- 647
Db      900 SLRSREIIECHAYLVPLVIRLLMRVVRKLKGLAKRKXSI---CHRKSILSFIINGLD 956

```

```

Qy 648 -----HPVSSLSOMISAEQVNAAGYVNLNNSSEOPQGEYVTLAFLITTLV 695
Db 957 VVEPLFPFALLIKPQIYK-----TDGPALFMTSDKSIDEPADALLEFTIDN 1008
Qy 696 KQGLSTOSQGLVPCVFMFLKEMLPYHAKWRYNHGVREQ-----IGCLILEL- 743
Db 1009 IANLSMKKRYGF-----LHVIEDII-----GVFDELHIREPFLDLGVCVRLLES 1053
Qy 744 -----IHALNLCHETDLHSSHTPSLOFLCISLATEAGQIVINIMGIVTIDMMAQ 799
Db 1054 CTSLSHNLNLGL-PSDQHN-----CSTSSNSLGE-----DSVPTNQ 1088
Qy 800 PRSDAEQO--GQGLILIKTVLAFSYTNVNLRLKPSNVVSPLEQALSQHAGNNLIA 857
Db 1089 TQINGTLQKDMBSLCKIISL-----ATVAPMSVYAC-LGNDAAAIRDAFLT 907
Qy 858 VLAKYIYKHDPALPRLAIQLIKRL-----ATVAPMSVYAC-LGNDAAAIRDAFLT 907
Db 1112 VLANKYEDHEFSSDLMDREFSAVKPLVDKFOEASSEKPSLLSCFLMSANNKLVALLY 1171
Qy 908 RLQSKIEDMRKIMLEFLYVAVERQPLIELFLNLKXDGSDGSKERSLGMSCLAHL 967
Db 1172 RKESLVPD-----IFSLISVNSASE-----AVIYCVL 1198
Qy 968 ELIDSQOQDRYWCPELHRAAIAFLHALMODRDSAMLVLRTPKFWENLTSPLFGT--- 1024
Db 1199 KFEV-----LTSLDNFENDBNSAQVLLSNIKVLMDSNCCFLGSGNA 1242
Qy 1025 -----LSPSETSEPSILETCALMKIICLEIYVYVKGSLDOSLKDPLKFSIEKRAYW 1079
Db 1243 IKRKLIKPSGET-----VIRILEFLPKYISEALIKQFODILLF-LENKTONS 1290
Qy 1080 SGYVKSILAV--HVAETEGSSCTSLLEYQVLVSAWMLLIITTHADIMHLDVSVRQLF 1137
Db 1291 DRYEALQVITQITPILHGSTA-----KILSAVSPLTISB-----LDMERLICDL 1337
Qy 1138 LDVLDTGALLLVPAVNCRLGSMKCTLLILRLQMK--RELGSVDEILGPLETEILE-- 1193
Db 1338 LDALVASPASLSLVAK-----LRLQMLATSTLGLMDH-----DALINAY 1376
Qy 1194 GVLDOOQUMERTAKVSAFTVYQMKEMKYSIDIPQVSLVANC--ETLOEVIYALF- 1250
Db 1377 GIIND-----FFRSYQV-----EHALLILSHCVADMSEETTFWFS 1413
Qy 1251 -----DQTRHSIAL--GSATEDKQSMETDPCSRSHRDQDGVCLGHLAKCEV 1306
Db 1414 AYSSLSLSTVDSAHILCOBNSSEQLSVKRTDSCWITSCIQTRAKKFLKHMADAM-- 1470
Qy 1301 DEDGD-----SWLQVTRRLPI-LPTL--LTTLEVSLRMKQNLHFTTEATLHLTLARTQ 1352
Db 1471 --DGLSVITKGMIKLHQMWLKLPEVSNLKSMLVLCNDEGVNFPD----- 1514
Qy 1353 GATAVAGAGITQISICPLISLYQVSTNGTAQTPASRSLSLAPSV---PGVYRLSLMSLM 1408
Db 1515 -----NITDSVY-----RRKRYALSMFRNVIVSNKFSBIT 1545
Qy 1409 EQLLKTLEYNFLPEALDFVGVHOERTLOCLNAVTVQS-----LACL-----EBA 1453
Db 1546 EKVEPRLEFPNMLYDEKSGAKEMKRA--CIFTIASVSGQMKSYALLICFPGASRSP 1603
Qy 1454 DHTVGFILQLSNFMKEWHFLPQLMRDIQVNLGYL-----CQAC--TSLILH 1497
Db 1604 DKQKLFIRLISILDKFHFSEVPNKEPKESLGGVSDMDITDTVNKEIQCLKVVLPK 1663
Qy 1498 SRKMLQHLQMN-----GDGLPSAVAVQRPSPASASAPSSSKQAPAD 1542
Db 1664 IQKLNDSSEKVNWNISLAALKLKLPGDVMDLYLPTIYHISNPLKSHLESIRDEARS 1723
Qy 1543 TEAS--EQOALHTVOYGLIKLISKTALRHFTPDVQCILDDQSILDAEYVFLALSFTY 1600
Db 1724 ALATCMLGLEYLOF-ILKVQSTLR--RGVEHLVGLYTL-----NFIISKCLIS 1771

```

```

Qy 1601 PTFDEVPAPSGTLLATVAVNLMLGELDKKKEPLTQVAGLSQAEGRITLSL--LMFT 1658
Db 1772 PV-AGRIDYCLELDSLSTIE--NDILGDVAEQK-----VEIASKMKETRRKSSRESLKV 1824
Qy 1659 MENCFFYLLIQAMRYLDRPAVHPBDKQRMKQELSSSELSTLSSISRYFRGAPSSPA 1715
Db 1825 AONVTFK--SYALKLAPVTAH-----LKKHITPVNKGLEMLNLQIATAGISBNPS 1873

RESULT 36
US-10-369-493-2224
; Sequence 2224, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 2224
; LENGTH: 1557
; TYPE: PRT
; ORGANISM: Schizosaccharomyces pombe
US-10-369-493-2224

Query March 1.7%; Score 155; DB 14; Length 1557;
Beat Local Similarity 16.2%; Pred. No. 0.0021;
Matches 303; Conservative 281; Mismatches 590; Indels 696; Gaps 83;

Qy 112 VKTVLQDEROSQALLIKIADYYEERTCIDLRLHLLTYFQDERHPRYVEADCVDLK 171
Db 2 IKLVDAEKQCN-----VQDASFERTWL-----WIPDKESVYKAW-----IVE 42
Qy 172 ELVSKTRQPE-----ELYTEAPFTWETHGNLMTEROVSRRFVOCLEBQML-- 218
Db 43 DLGSKYFVKLERGSEIIVDFPAEKVNPPEKPMVDMMA-----LTCLEPSSVNN 94
Qy 219 -----LEIFLYAVFEMAPSDLVL-----TKMKF----- 244
Db 95 LTQRYEKDLIYVYSGFLVAVNYPYCHLPYGDVVRKYQSKQKREKPHIFGDAAYRS 154
Qy 245 -----EOGFGSRQTRNHLVDETMDFVDRIGYFSAALLVBSGMDIESLHKA 290
Db 155 LEBRINQSLVTVGESAGKETTKKVIQ-----YLSVNDASTSDSQLEKLI 203
Qy 291 LDBRRELHQAOGLICQMDC-----LMLTFQDIPHHAVLALMALR--HTLNPEFS 343
Db 204 LETNPVLEAFGNAQVYRNNNSRFGKFIREFSNNQSIWGANIDWVLEKSRVYHPSNE 263
Qy 344 -----SVYRKUGTAIQ-----LNVQYVTRLLQSLASGNDCTTSTACGCVGL 389
Db 264 RYHNVFQILRGADGSLLESFLDRYVDHYSYKNGIKHI--NGVDDGKFPQKLC-FGLR 320
Qy 390 SFVLTSLEHTL-----GNQODIIDTACEVLAD-----PSLPELFMGTEPTSG 432
Db 321 TLGFDNNEHISLPLIASILHGN-----IEVASDRSGQARFSLTG----- 362
Qy 433 LGIILDSVCGMPHILS-PLQLLRALVSGSKSTAKVYSFLDK-----MSFINELY 482
Db 363 -----IDQLC-----HLLEIPVDGFVNAALHFKSAGREWIYTAETRBQVHTTQSLAKGLY 414
Qy 483 KHKPHVDISHEDGTLWRQTPKLLYPLGGQTNLRIPQGTGVQVMDRAVLVWEXYSYS 542
Db 415 ERNPAHLVVKRLNQTMYTYSSEH-----DGFIG--VLDIAGEFI--FTFNS 455

```



```

Db 794 DQPAKKEKTEVSPILNV-----ECTENNALL-----LISTETPISISS 834
Qy 401 LGNODDIDTACEVLAADPSLPFENGTEPTSGIILDSVCGM-----PPLISPLQILR 456
Db 835 SRSIDLFISKIOMELASAGIPNVY-----VPLVINGLFGILNRRSYLMNPVLEICIA 886
Qy 457 ALVSGKSTAKKVVSYFL-----DKMSFYNELYKHKPHDVISHEBGTQTRQTPKLVLPLGQOT 513
Db 887 VLIS-----LHFLRWSDSLVAVLERQ-----TIF--DIPSNLH----- 918
Qy 514 NLRIPOGVQVMDLDRAYLVWMEYSYSWTLFTCEI--EMLLHVSTADVIQHCORVXPI 572
Db 919 -----GSVNGALFPOPAGLV-----DOFKLFVYHASDSFSTVITALLQALQKIPV 966
Qy 573 ID----- 574
Db 967 IEPRSRQPLPLFKELGYPDVLSVGLFDSHACKGEMKAILKEMWNLKLMKNPKSPCYG 1026
Qy 575 -----LVHKVISTD-----LSIADCL-----LPITSRIYML-----ORLTT--VI 608
Db 1027 QELKOVLRHRLLENDTEIQKRVADCLIMDDYILFVEHRLNLSKRLREBELTWSL 1086
Qy 609 SPPVDVIVASC-----VNCITVLAARNPAKYWTDLRHTGFPFVA----- 647
Db 1087 SRESEIIIECHRAVYLVPIVIRLIMPRVYKLGASRKKASI-----CHRSIISFIAGLDV 1143
Qy 648 -----HPVSLSQMISAEKNAGGYNLMSQPOGEYVTAIFRLITLTVYG 697
Db 1144 ELPLFALLIKPLQIVKK-----TDGPNALFWTSDKVSIDEQADALLEYFLDNIA 1195
Qy 698 QLGSTOSOGVPCVWFVKEMLPYHKRYSNHSVREO-----IGCLILEL----- 743
Db 1196 NLSMKKKYGF-----LHVIEDII-----GPFDELHRRPLDILVCGVALLBECT 1240
Qy 744 -----HAIILNLCHETDLHSHTPSLOPLCISLAYEAGQTVINIGIVDTIDYMAAQR 801
Db 1241 SSLHNLNGL--PSDOHN-----CSTSNLSIGL-----DSVPTQOTQ 1275
Qy 802 SDGAEQO--GOGQOLIKTVKLAFTVNNVIRLKPSPNVSPDLQALSHGAGNNLAVL 859
Db 1276 INGLINOLKDMRSLCKIISL----- 1298
Qy 860 AKYIYHKDPAFLPLAIOQLKRL-----ATVAPNSVYAC--LGNDAAIRDAFLRL 909
Db 1299 NKYEHEPSSDLMPDFSAVXPVLDKFOZAASSKPSLSLSCFLMANNKVALLYRX 1358
Qy 910 QSKIEDMKIKMILEFLTVAVETQPGILEFLNLEVDGSGSKSPSLGMSCHAVTEL 969
Db 1359 ESLVPD-----IFSISVNSASE-----AVIYCVLKF 1385
Qy 970 IDSQOQDRYWCPLLRRAIAFLHALMODRDSAMLVLRTPKFMENLTSPLFGT----- 1024
Db 1386 VEN-----LSLONERFEDENSQORVULSNIKYLMBSMCLFSSDNAIK 1429
Qy 1025 -----LSPPEETSEPSILETCALIMKIICLEYVYVVGSLDQSLKDTLKKFSIEKFPAYWSG 1081
Db 1430 RKLKSPGET-----VIRILEPLPKYISEABELAQFVILILLF--LENKTONSDV 1477
Qy 1082 YKXSLAV--HYAETEGSSCTSLLEYOMLVSAMRMLLIATTHADIMHLSVVAROLF 1139
Db 1478 RVEALQVQNIITPLIGHSTA-----KILSAVSPLYISAE-----LDMRLICDL 1524
Qy 1140 VLDTGKALLLVASVNCILRSGMKCTLLILLRQMK--RELGSVDEILGPLTEILR--GV 1195
Db 1525 ALVASDASLSLVAK-----LLRQLANSTLIGMDH-----DALINAYGI 1563
Qy 1196 LQADQOLMEKTKAVFAFITVLQMKEMKVSIDIPQSOVLVNC--ETLOEVIALF-- 1250
Db 1564 INTD-----FFRSVQV-----EHALILSHCVHMSSSETTFMFSAY 1600
Qy 1251 -----DQTHSLAL--GSATEDDSMETDSCSRHRDQDGVYGLHAKELCEVDE 1302
Db 1601 SSSLFVDFSAHILCOBGSNEEOLSVMNKNTDSCTKCIORTAKKFLKHMADAM----- 1655

```

```

Qy 1303 DGD-----SWLOTRRLPT--LPTL--LTTLEVSIRKONLHFTETATLHLLTTARTQOGA 1354
Db 1656 DGLSVIKGWIKLIHQVNLKLPVSNLKSIMVLCNEDGEVNFED----- 1699
Qy 1355 TAVAGATIOSICLPLISVYQLSTNGTAQTPASRSKSLDAPSW-----PGYRLSMSIMEQ 1410
Db 1700 -----NITDSVL-----RKRVALISMFRNVI SVNKSSEPIITEK 1732
Qy 1411 LKTLRYNPLPEALDFVGHQERTLOCLNARVYOS-----LACL-----BEADH 1455
Db 1733 VFMRLPFNMLYDEKEGAEMHKA--CIETIASVSGQMKRSYALLIRCFMGASRPDK 1790
Qy 1456 TUGFILOLSNPMKEMHNLQMLRDIOVNLGYL-----COAC--TSLHSR 1499
Db 1791 QKLFIRLISILDKFHSSEVPNHNKPEPSISGVSMDITDPTDVNKEIOTCLYKVVLPKIQ 1850
Qy 1500 KMLQHYLQNRX-----GDGLPSAAGVORPSPASAPSSKOPADTE 1544
Db 1851 KLANSDEKYNVNI SLAALKLLPLRQDVMDLYLPTVIRHISNPLSHLESIDEARSAL 1910
Qy 1545 AS--EQQALHTVOYGLKILSKTILALRHPTPDVCOILLDQSLDLAEYNFLPALSTPTPT 1602
Db 1911 ATCLKELGLEYLQF--ILKVLQSTLR--RGYELHVLGYTL-----NFLSKCLSSPV 1958
Qy 1603 FDBEVAPSFGLIATVAVNALMTEGELDKKEPILTOAVGLSTOABGRITLKSLL--LMTME 1660
Db 1959 -AGKIDYCLBLLSVIE--NDIIGDVAEOKE-----VEKASMKKETRKRKSFSLKVAQ 2011
Qy 1661 NCFYLISQARYLRBPVAPRDKQMKQELSELSTLSLRYPFGAPSSPA 1715
Db 2012 NVTFK--SVALKLAPVTAH-----LKGIITVNVKGLKEMLOHATGIESNPS 2058

```

```

RESULT 38
US-10-723--860-2568
; Sequence 2568, Application US/10723860
; Publication No. US20040253606A1
; GENERAL INFORMATION:
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsburg, Wendy M.
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &
; TITLE OF INVENTION: Methods for Screening for Soft Tissue Sarcoma Modulators
; FILE REFERENCE: 05882, 0193, NPLUS01
; CURRENT APPLICATION NUMBER: US/10/723, 860
; CURRENT FILING DATE: 2003-11-26
; PRIOR APPLICATION NUMBER: 60/429, 739
; PRIOR FILING DATE: 2002-11-26
; NUMBER OF SEQ ID NOS: 8393
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2568
; LENGTH: 3830
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-723--860-2568

```

```

Query Match 1.7%; Score 153; DB 17; Length 3830;
Best Local Similarity 17.2%; Pred. No. 0.015;
Matches 331; Conservative 285; Mismatches 658; Indels 654; Gaps 88;

Qy 54 YKPPSSSAEKVVA-----NKDVASPLKEGLRI--SKFLG-----LDDEOSVOLLOCYLO-- 102
Db 862 YDHIOQVRAEMALMQLRTLRNPADSLSHAYAYVLGFGGSKNKMLESQKLRHVTVTEVQG 921
Qy 103 -----EDYGRTRDSVKTVLQDERGQALILKID--YYVEERTCIARCVLHLITTYQDER 155
Db 922 PSITVEFSOCKASLQPMKEALFETALDCKSANTEBYRQRAWEVVKCFVLVAMMSJEDNK 981
Qy 156 HP-YR--VEYADCVDKLEKLVSKYRQOFEBLYXT--BAFWETHGNLMTERQVSRWVQCL 212
Db 982 HALYQLLAHNPFTEKTIPIVITISHR-----YKAQTPARKTFEOLATGAFMSA-VIKOL 1034

```



US-10-437-963-108860

Query Match 1.7%; Score 152.5; DB 16; Length 2159;

Beet Local Similarity 18.7%; Pred. No. 0.062; Matches 328; Conservative 243; Mismatches 574; Indels 609; Gaps 82;

227 AYFENAP---SDLLVTFKFEQ-----FGSRQTRHLVDETMDFV-----DRI 269  
 213 AIFEVSSGSLSDHDKMKIFVTEGVPTLMIDLKRSKH-QDKVGEFTGALRNLCGKD 271  
 270 GYFSALLIVEGMDISLHKCALDRRELHOFADGLICQDMCLMTFGD-IPHHAPVL 328  
 272 GYWRANLRAAGVEIIT---GLISSKTTTSSNAASLAR---LVSAGDSIPKIIDAGA 324  
 329 AMALIRHLNPE-----ETSSVVRK---IGGAIQIOLNFOYITRLQ 367  
 325 VKALIR-LNKRNDISVRESAADALEALSSKSIKKAVDAGGIPV-----LIG 373  
 368 SLASGNDCTTSTACMCYGLSFVLTSLHTLGNQODIIDTACEVLADPSLPBLFWGT 427  
 374 AVVAASKECMGDTG-----HSLQSHAVHALSN-----IC----- 403  
 428 EPTSLGIILDSVCMFPHLSPILLQILRALVSGSTAKKYSPFLDKKSPFNLKXKPH 487  
 404 GGTVSLILLYGELC-QVPSPPVPLADILGAL---AYTLWVFSGTGKSF----- 448  
 488 DVISHEDGTWRROTPEKLLYPLGGQTNLRIPQGTVGVMDDBRATLVWEXSYSWTLFT 547  
 449 DPIEIE-----NILI-----VLK-----SYSSNLVLD 471  
 548 CEIEMLHVSTADY---IOHCQVKKPIIDLVHKVISTDLSIADCLPITSRIYMLQRL 604  
 472 RILEMLASLYGNACSLGRHNSNAKVLVGLI----- 503  
 605 TTIVSPRPDYVASCNCLTVLAARPAKWTDIRTGFLPVAHPVSSLQMSIAEGMNA 664  
 504 -TMA--ADVQKLVHALTSLCS--DGIGIMDLGR-----EGTOLLISFLG--- 546  
 665 GGYGMLNNSRPOGEVGTIAFLRLITTLVKQGSTOSQGLVPCVMFLKEMLPYHK 724  
 547 -----LSRHOQ-EYAVSL--LALISDEVDDSKALTAAGGIPVLQLET----- 589  
 725 WRYNSHGVRQEGCILELHAILNL-CHEIDL----- 756  
 590 ---SQXAKE-----DAHILMWLCHSDDISACVESAGAVLALLMLKSGSPHQEA 639  
 757 -----HSHPTSLOFLCISLAYTEA-GQTYINIMGIGVDTIMVMAAQRSD 803  
 640 SAKALKITIRSAADSTINQLRALISDSLSTKHAHITVLGHVLAWSQORDLVONGAPANK 699  
 804 G-----AEGQGGOLLIKTVLAFSVTNNVIRLRKPSNVSPLEQALSQGH 851  
 700 GLRSLIDLESSNETQGAATVADI---FSTRQDIDDIIGTBEIITPOCKKLITS--- 752  
 852 GNNLIATLAKTYHKHDPALPRALIQILKRLATVAWSVVAACGNDAAIRDAFLTRIQS 911  
 753 GNOVIAT-----QSRALGALS-HSANMLKMKKSCIAEGVQTLIE 793  
 912 KIEEMRIK-----VMILEFLT---VANETOPGLIELFLNLEVDS--DSKEKESLQMS 961  
 794 MSKSPSIDAAETTIALANPLSDAHIAEALDGNIVALTIVLEGSLEGKISASRSILQ 853  
 962 CLH-AVLELIDSOQODRYWCPPLHRAAIFALHLMODRRDA-----MLVLRTPK 1012  
 854 LLNGPLNEVLPDVSQ---CYFIH---ALIVCLGSINENNTNDPLNVLAMARXK 905  
 1013 FWEMLTSLFGLTSPSETSE-----PSLFTCALIMKTIICETIYVVGSLDQS 1062  
 906 EGAFHSSPLMGAFLDVPESLEPLVACISVGLPPIQDKAIQITIALCQDQPSILGEHLNRS 965  
 1063 LKOTLKKRSIEKRFAYMSGVYKSLAVHYAETE-----GSSCTSLLEQMVSMARMLLI 1116  
 966 -----QGTIASLASRVISBSTMEIRIGSALT-----LISMR----- 997

1117 IATTHADIMHLTDSVVRQELFLVDYDSTKALL-LVPAVNCIRLGSMMKCTLLILLRQWK 1175  
 998 ---HS-----REHSIDVTEASGHLKNLISISIDMMKODSAPTSL---DIWVK 1039  
 1176 ---RELGSVD-EILGLFTEILSGVLQADQOLMEKTKAFSAFTYVLQW-KEMKVS DIP 1229  
 1040 PYRPNLSLYNDKQVIG-----VSGSGKVLBEYALWALNLSICSHSLSKULTVND-- 1088  
 1230 QYSQVLNVCETLOEYVIALPDQTRHSLAIGSATDEKOSMETDDCSRRHRDQDGVCL 1289  
 1089 ---LGVETTISDKLASV-----TTNQODSMVLVOSP----- 1116  
 1290 GLHAKELCEVDEGDSWLVQTRPLPILPTLLTLEVSLSMKONKHFTEATL-HILLTLA 1348  
 1117 -----IMRTISLASLSKDKIT-----DKTFAQSILASLVSTGS 1151  
 1349 RTQOGATAVAGA-----GITQSICLPLLSV---YOLSTNGT-----AQT 1385  
 1152 RSIGLAIANSAGAVGTIAMIGQESTWPNLVMAABEFLKADNPISKILRSLFELEDVRTS 1211  
 1386 SASRKS-----DAPSNPQYRUSMSLMEQLTKRYNPLPEALDFVGHQERTLOC 1437  
 1212 ATARRSIPLLVDLKEPDRQAPLVALHLLTQLABGSEFTNKV--AAABAGV----- 1261  
 1438 LNAVRTVQSILCEADHTVGFILQLSNFMKEWHFHLPOLMRDIOVNLGYLCOACTSLH 1497  
 1262 LDALITKLSLSPQDSTETITINILIRIITYPDDLYHSSSISTNQVAVARLGRSRNRLN 1321  
 1498 SRKKLOHYLONKKGDLPSAQAQRVORPPSAAASPSKOPADTEASEQOAL----- 1551  
 1322 AARTQLQLPSEN-----IRDTEVAMQAIPLPLDMLESGETEGQOALGALIKL 1370  
 1552 -----HTVOYGLKILIS-KTLAARHTPVOQIILDDQSLDLA----- 1589  
 1371 SSGNISKASALFVEGTTLESILKILSFSSLSLEKMDAQCYLEENSTIRASPIASEC 1430  
 1590 YNPLFALFTPTPEDESVAPSFGLTATVVALNMLGELDKKEPTLQAVGLSTOAGETR 1649  
 1431 LQPLISLMTSGSTFVVEPAR-----ALNRLLDETYNME-----IAATSEVVD 1473  
 1650 TLKSLMFT---MENCFYLLISQAMRYLRDPAVHPDRQKRMQEL----- 1691  
 1474 LTVSPVGTNHQJSEACIGALIKL-----KDRNCTLEWVKAGIIEHVLDMIL 1522  
 1692 ---SSELSTLSLSKRFRRGAPSPR-----TGVLPSPQCKSTLSLKASP 1734  
 1523 DVPVSVSSSTIAELLRIITN--NSGIAKSSAAAMSHSALQALVNLERPO--SLAALKLSP 1578  
 1735 ESQ--EPLIQVOA 1746  
 1579 -SQIIEPLISFLS 1591

## RESULT 40

US-10-408-765A-214

Sequence 214, Application US/10408765A

Publication No. US20040101874A1

GENERAL INFORMATION:

APPLICANT: Ghosh, Soumitra S.

APPLICANT: Fahy, Eoin D.

APPLICANT: Zhang, Bing

APPLICANT: Gibson, Bradford W.

APPLICANT: Taylor, Steven W.

APPLICANT: Glenn, Gary M.

APPLICANT: Wainock, Dale R.

TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION

FILE REFERENCE: 660088.465

CURRENT APPLICATION NUMBER: US/10/408,765A

NUMBER OF SEQ ID NOS: 3077

SOFTWARE: FastSeq for Windows Version 4.0

```

; SEQ ID NO 214
; LENGTH: 2412
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-408-765A-214

```

Query Match	1.7%;	Score 152.5;	DB 16;	Length 2412;
Best Local Similarity	18.5%;	Pred. No. 0.0075;		
Matches 397;	Conservative 234;	Mismatches 765;	Indels 689;	Gaps 100;

Qy	33	LSQIAEAEINRHRRLLEGJS-YKPPSPS-SAEKVANKOVA-----SPLKIGLRISK	84
Db	26	LASVTLDSQYAMDIYKGLAGHKXSPRLMBDAVALALNRLNAQCCSDSMAMEBLTYKHLFA	85
Qy	85	FLGLDEBGSVOLLQCYLQEDYRGRTDSTVKTVLDEROSQALLIKIADYYEERTCILRCV	144
Db	86	ILGGSEGRKLTVAQAQ---KMSVLGSGISVSHHVVGSSQVLANGIIVALEF-----	1311
Qy	145	LHLTLFYQDERHPRYREYADCV-----DLBLEKVLKRYKNOQPELYKTAFTPMETHGNLM	199
Db	132	---IPFLOQEEVNEGTLVHAVSVLALMCMNPFMEVPKKLTWFKFAESLKST-----	1801
Qy	200	TERQVSAMEVQCL-----REQSMLEIIFUYAFAEMAPSDPLVLTWFKEQGFSRQTNR	255
Db	181	--SAVRHAYLOCMASRYGDTLLQAL-----DLPLLLQTYEKA-ASQSTQV	224
Qy	256	HLVDETMDFVDRIQYFSALIVE-----GMDIESLHKCALDDRRELAHQ-----FA	3011
Db	225	PLITE-----GVAAALMLLLKLSVADSQAERKUSSFQGLVDERKQVFTSEKFLVMA	2751
Qy	302	ODGLICQMDCLMTFGDI.PHH-----APVLLA--WALLRHTLNEEETSSVY	346
Db	276	SEDALCTVHLALTERFLDPRPHRLTGKNOQYHRALVAVLLSRTHMR-----QAQQTV	329
Qy	347	RK-----IGGTAIQLVNFOYLTRLLOS-----LASGNDCTSTACMCVYGLSLPVL	394
Db	330	RKLSSLGCFKLAHGLBELKTVLSHKVLPLEALVTTDAGEVTGAKAYPRPVLOEALC	369
Qy	395	SLEHLTLGNQODIIDT---ACEVLA---DPSLPFLFWGTEPTSGLGIIILDSVCGMPEHL-	447
Db	390	VIS-GVPRGLKGVDTDEQAQEMLLISHNPSTLAVQSGLWPRALLAMKIDPEAFITRRHD	448
Qy	448	-----LSPLLQILRALVSGKSTAKKATYSFLDKNSFTNELYKHKRPHVYISHEGDTLMR	499
Db	449	QIIPMTTOSPLNQ-----SSNMANGSLSVLSPDRVLPOLISTITA	489
Qy	500	R-QTPKLYPLGGQNTLRI.PQGVGOVMDLDRAYLVRWE-----YGYSSW	543
Db	490	SVQNPAL--RLVTRREFALMQTPAGE--LYDKSIIOQAQDSIKKANMKRKNKAYSPKEQ	545
Qy	544	TLFTEIEIMLHVSTADVIOHCQVRKPIIDLVHAKTISTDLADCLPITSRITMLOOR	603
Db	546	II---ELBL-----KEIKKKKKGKEEVOLTSK---QKEMLOAQDLREAQVRRLOE	591
Qy	604	LTTVISPPVDVIASCVNCLTVLAARNPAKV--WTDLRHTGFLRFVHAPVS-----S	652
Db	592	LD-----GELEAALGLDIIILAKRPSGLTYQIVPLVDSFRLKSPILAARPIKPFLS	644
Qy	653	LSQMTSABEGMANGG--YGNULMNSBP-----QGEVGYI--AFRLITLVYKGL	699
Db	645	LAAVCMPSRLLKGLTVLHVSHVTLRLKRECVCLDKSWQCEBLSAVAKVAMLLHHTTITSRV	704
Qy	700	G-----STQSGOLV-PCVMFVLKEMLPYHKWRVNSHGVRQEOIGCLILELHAIYN	749
Db	705	GKGEBCGAAPLSAPSLVPRPKMVLTEM-----PHHSEEE-----EMMAQIIG	749
Qy	750	-LCHETDLHSS-HTPS-----LOPLCICSLAYTEAQTVINIMIGIADVTDIMVMA	797
Db	750	ILTVQAQLNASNPTRPGRVDENGPELLPRVAMRLTLTVIGTQSPRLQVLAASDTLTLTCA	809
Qy	798	AQPRBDG---ABGQGGQGLIKTVKLAFSVTNNVITLKPPSNVSPLEOALSQGHAGNN	854
Db	810	SSSGDGCAPAEQOEVDVLLCALQSCBASVRETVLGLMELHNVLLPAPDTERKNGL--N	866

QY	855	LIATVAKTYIYKHDPBALPRLAIOLIKRLATAVAPMSVACIGNDA----	AAIRDA----	FLT	907
Db	867	LIRKLIVVYKFPKEE-EIRKLABERLMSMGDLDOEDLCSLIDDDVIYHEAAVROGAEKLS			925
QY	908	RLQSKIEDMRIRKVN--ILFELTVAVETOPGLIEFLMLEVKGSDGSKXESLGMWSCJHA			965
Db	926	QAVARYGQOALEVNGRLMEIYQOEKLYRPPVDAI-LGRVISESPDQWEARSGJLALANK			984
QY	966	VLELIDSGOOD---RYMCPRLIHHAAALFHALMODR-----	DSANMLVATKPKFW		1014
Db	985	LSQYLDSSQVKPLQFPVFDALN-----	DRIPDVRCMLDAAALTLNTHG--		1030
QY	1015	ENLTS--PLF-----	GTLSPPSESTSPSILFETCALIMKIT-		1047
Db	1031	ENWVSLPVEEFELKNAPNDASVAVRQSVVVLMSGSLAKHLDKSDPKKPIVAKLIJALS			1099
QY	1048	-----CL-EIYVVK-----	GSIDSLKDTL--	KKFSIEKRPAY--	MSGVYSLA 1087
Db	1091	TPSOQOESVASCLPPLVPAIKEDAGMIQRLMOQLLESIDKARXKGAAYAGIYVKGLG			1155
QY	1088	V-----	HVAETSSGTSILEYOMVYSANRMLIITATVADIMHL		1122
Db	1151	ILSLKQEMMALTDALIODKKNFRFRBQ----	ALPAPMLCTMIGKLEPYVYH-----		1200
QY	1128	TDSVVRRLFLVDVGTKALL-----	LVPASVNCILRLGSMCTL--	LLILRQ--	WKREL 1178
Db	1201	-----VLPHLLLCFQDGNQYREAADDCAKAVMSNLISANGVLTPLSLAALBESBWRKA			1256
QY	1179	GSVDEILG-----	PLTIEEG----	VLQAOOQMLEK-----	1205
Db	1257	GSV-ELIGAMAYCAPKOLSSCLPNIVPKLTIEVLTDSHVKKYQAKAOQALRQGSIVIRNBEI			1315
QY	1206	-----	TKAVFSAFITVLQMKKEKVSVDIPQYSOLVAVNCETLQE-----		1244
Db	1316	LAIAPVLLDALTDPSRKTKQKLOTLDBTKRPHFIIDABSLAIMPIYQRAFDGRSTDRKM			1375
QY	1245	-----EVIALFDOTRHSIALGASATEDKSMETDCCSRSHRDQD--	GVCYLG----		1291
Db	1376	AAQIIGMYSITLDDKDLAPLYPSVTPGIKASILDVPEVPRVTSAKALGAMVKGMGESCFE			1439
QY	1292	-----	HLAKELCEVDEDS-----		D 1305
Db	1436	DLEPMLMETLTYEYSSVDRSAGAGLAEWAGLGEVKELEKMPETIVATASKVDIAFHYRD			1495
QY	1306	SWLOVTRRL-----	PIPLTLLTTL-----	EVSLRMKONITPTEATLHL	1344
Db	1496	GYYMMFVNLPIFDGDKFTPYVGPIIPCILKALADBNFVRDTALRAGRGVISMVAFYAIA			1555
QY	1345	LTLARTOOG-----	ATAVAG-----	AGITOSICLPILSVOLSTNGRQTPSA	1387
Db	1556	LLLPOLBEGFLPDDIMRIRFSSVOLIGDLFPHISGVTGMMTETAS--	EDDNFGAQSOKA		1613
QY	1388	SRKSIDAPSMGVGYRLSMSLMEQLKTLRYNPLFEALDPFGVHOBERTLOQANAVRTVQSL			1447
Db	1614	IITLH-----	GVRR-----		RNRVLAGLYMGRSDIQL 1633
QY	1448	ACLEBAHTVGFILQLSNFKEMKHFHLPOLMRDI-----	QVNLGYLCOACTS-----		1494
Db	1640	VVROASLHVWKIVVS-----	NPRJTLREIPLTFLGILLGFIASTCAKDKRTIART		1689
QY	1495	-----	LHSHKRMLOHYLONKRGD-----	GLPBAVAVORVPRP--	SAASAA 1532
Db	1690	LGDVLKKEGKILPEIIPITEEGRSQSDRQGVCIGL--	SEIMKSTSRDAVLVYFSESLV		1748
QY	1533	PSSSKOPADMEASEQOALHTVOYGLKLISKTIA--	ALRHFTPDVCOILIDOSIDLAEY		1590
Db	1749	PTAKKALCDPLBEVREBAAKTFEQ-----	LHSTIGHQALIEDILPFLKQLDDE--	EVESE-	1800
QY	1591	NFLPAL-----	SFTTP-----	TPSEVAPB-----	FGLLATV 1618
Db	1801	---	FALGOLKOVAMIKSRVVLVPLVPKLITPPVPRVTLAVLAFSSVAGDALTRHGLVILPAV		1857

